LIBRARY AND INFORMATION SCIENCE TEXT SERIES

FOURTH EDITION

THE ORGANIZATION OF INFORMATION

Daniel N. Joudrey and Arlene G. Taylor,
with the assistance of Katherine M. Wisser

The Organization of Information

Recent Titles in
Library and Information Science Text Series

Internet Technologies and Information Services, Second Edition
Joseph B. Miller

Library Programs and Services: The Fundamentals, Eighth Edition
G. Edward Evans, Margaret Zarnosky Saponaro, Carol Sinwell, and Holland Christie

The Complete Guide to Acquisitions Management, Second Edition
Frances C. Wilkinson, Linda K. Lewis, and Rebecca L. Lubas

Introduction to Cataloging and Classification, Eleventh Edition
Daniel N. Joudrey, Arlene G. Taylor, and David P. Miller

The Collection Program in Schools: Concepts and Practices, Sixth Edition
Marcia A. Mardis

Libraries in the Information Age: An Introduction and Career Exploration, Third Edition
Denise K. Fourie and Nancy E. Loe

Reference and Information Services: An Introduction, Fifth Edition
Linda C. Smith and Melissa A. Wong, Editors

Guide to the Library of Congress Classification, Sixth Edition
Lois Mai Chan, Sheila S. Intner, and Jean Weihs

Research Methods in Library and Information Science, Sixth Edition
Lynn Silipigni Connaway and Marie L. Radford

Information Services to Diverse Populations: Developing Culturally Competent Library Professionals
Nicole A. Cooke

The School Librarian as Curriculum Leader
Jody K. Howard

The School Library Manager: Surviving and Thriving, Sixth Edition
Blanche Woolls and Sharon Coatney

THE ORGANIZATION OF INFORMATION

Fourth Edition

Daniel N. Joudrey and Arlene G. Taylor

with the assistance of Katherine M. Wisser

Library and Information Science Text Series

LIBRARIES UNLIMITED™

An Imprint of ABC-CLIO, LLC

Santa Barbara, California • Denver, Colorado

Library of Congress Cataloging-in-Publication Data

Names: Joudrey, Daniel N., author. | Taylor, Arlene G., 1941– author. | Wisser, Katherine M., author.
Title: The organization of information / Daniel N. Joudrey and Arlene G. Taylor ; with the assistance of Katherine M. Wisser.
Description: Fourth edition. | Santa Barbara, California : Libraries Unlimited, [2018] | Series: Library and information science text series | Significant expansion of: The organization of information. 3rd ed. / Arlene G. Taylor and Daniel N. Joudrey. | Includes bibliographical references and index.
Identifiers: LCCN 2017026113 (print) | LCCN 2017048240 (ebook) | ISBN 9781440861291 (ebook) | ISBN 9781598848595 (hardcover : acid-free paper) | ISBN 9781598848588 (softcover : acid-free paper)
Subjects: LCSH: Information organization. | Metadata.
Classification: LCC Z666.5 (ebook) | LCC Z666.5 .T39 2018 (print) | DDC 025-dc23
LC record available at https://lccn.loc.gov/2017026113

ISBN: 978-1-59884-859-5 (hardcover)
978-1-59884-858-8 (paperback)
978-1-4408-6129-1 (ebook)

22 21 20 19 18 1 2 3 4 5

This book is also available as an eBook.

Libraries Unlimited
An Imprint of ABC-CLIO, LLC

ABC-CLIO, LLC
130 Cremona Drive, P.O. Box 1911
Santa Barbara, California 93116–1911
www.abc-clio.com

This book is printed on acid-free paper ∞
Manufactured in the United States of America

CONTENTS

PREFACE

The preface to the first edition of this book began as follows:

> As I began work on the ninth edition of *Introduction to Cataloging and Classification* I became more and more aware that another work was needed that would precede the cataloging text. Core courses in many schools of library and information science now include a course in organizing information. These courses typically cover much more than cataloging and classification. They discuss the concept of organization and its role in human endeavors; many kinds of retrieval tools, such as bibliographies, indexes, finding aids, catalogs, and other kinds of databases; encoding standards, such as MARC, SGML, various SGML DTDs, and XML; creation of metadata; all kinds of controlled vocabularies, including thesauri and ontologies, as well as subject heading lists; classification theory and methodology; arrangement and display of metadata records and physical information-bearing packages; and system design. *The Organization of Information* addresses this need, leaving *Introduction to Cataloging and Classification* as a textbook for courses devoted to the specifics of cataloging and classification.[1]

This statement is as true now as it was when it was written nearly 20 years ago. Changes have come in degree, however. More schools of library and information science now include organizing information as a core course.[2] Encoding standards, metadata standards, and systems continue to evolve, new ones have been developed, and old ones have disappeared. New information organizing concepts (or some would say old concepts with new names) have become useful additions to the discussion of organization of information—concepts such as linked data, tagging, and microdata.

The goal of the fourth edition remains to enable students, practicing librarians, and others interested in organizing information to understand the theory, principles, standards, and tools behind information organization in all types of environments. To this end, more figures and tables have been added to better illustrate the models, standards, and practices used in organizing information. Two textboxes have been added at the end of each chapter. One lists significant terms from the chapter and the other lists acronyms. All of these terms are defined in the glossary as well as in the chapter where they are listed.

As with the first three editions there is still more about libraries than about other types of environments. This is at least partly because the weight of history means that there are many centuries of accumulation of theory, principles, and practice behind organization of information in libraries, while archives have not quite two centuries of accumulation, and the other environments have only decades or, in some cases, less than a decade of history.

In order to accomplish the book's goal effectively, it was necessary to make some rather major changes in the structure of the work. The first three chapters continue to cover the same concepts as in the earlier editions but with extensive updating and some reordering. Chapter 1 looks at our basic human need to organize and how it is approached in various contexts. A new section on online settings has been written; it includes greatly revised subsections on the Internet, digital collections, and information architecture, as well as a new subsection on the Semantic Web. A discussion of personal information management also has been added. Chapter 2 is now the history chapter. It has been moved to an earlier position in this edition to encourage readers to ground themselves in an understanding of what has come before. In Chapter 2, we discuss the history of basic principles that have developed over the centuries. In this edition, there are amplifications of a few of the historical occurrences that were treated more briefly in earlier editions, particularly those associated with more recent history. Chapter 3 is concerned with the formats and functions of basic retrieval tools. The individual sections on catalogs, indexes, finding aids, museum registers, and search engines have been updated, and a new section reinforcing the need for specialized retrieval tools has been placed at the end of the chapter. Chapter 4, "Systems and System Design," (formerly Chapter 6) has been moved so that it follows directly after the retrieval tools chapter. The two chapters are related; they can be read together to get a better sense of the technology and system issues that underlie the profession's various retrieval tools.

After these initial chapters that provide background and context for the activities of information organization, the following chapters begin to discuss the processes associated with organizing information. Chapter 5 discusses the broad concept of *metadata*. It has been restructured and updated, and the section on metadata models has been expanded and revised to include the draft IFLA Library Reference Model (LRM) and brief introductions to Records in Contexts (currently being developed for archives) and the CIDOC Conceptual Reference Model (used in museums). In addition, the chapter includes expansions and updates to the sections covering the Resource Description Framework and the Dublin Core Abstract Model. "Encoding Standards," now Chapter 6, still deals with syntaxes used to format records, but it is expanded to introduce the ongoing Bibliographic Framework Initiative, a system being designed as a possible replacement for the MAchine-Readable Cataloging (MARC) format.

Chapters 7 and 8 continue discussion of metadata, covering resource description in Chapter 7, and dealing with access points and authority control in Chapter 8. The revisions to Chapter 7, "Resource Description," are significant. The chapter begins by examining points of convergence found among different organizing approaches; for example, libraries, museums, archives, indexing services, and others have common concerns, such as what a resource is called, who created it, when it was made available, and so on. These commonalities are addressed first, and then individual standards are described. The most significant change within the section on standards is the inclusion of *RDA: Resource Description & Access*, which was not yet available when the previous edition was published. Some metadata schemes that are no longer in use have been removed. Chapter 8, "Access and Authority Control," has been updated to more fully address RDA, IFLA standards (such as the *Statement of International Cataloguing Principles* and the LRM), and Encoded Archival Context–Corporate Bodies, Persons, and Families.

Subject approaches to organizing information are covered in Chapter 9, "Subject Analysis," Chapter 10, "Systems for Vocabulary Control," and Chapter 11, "Systems for Categorization." These are updated and rewritten from the third edition. It seems useful to expand on the challenges and concepts involved in determining the *aboutness* of an information resource; this enlarged treatment is covered in Chapter 9. Chapter 10 has been restructured so that it discusses, first, different types of controlled vocabularies (as identified in *ANSI/NISO Z39.19-2005 (R2010): Guidelines for the Construction, Format, and Management of Monolingual Controlled Vocabularies*). This is followed by a discussion of concepts involved in creating controlled vocabularies, which then flows into a section of applying controlled vocabularies, including descriptions of several of the most used existing vocabularies. In this edition, additional vocabularies, such as the *Library of Congress Genre/Form Terms for Library and Archival Materials* and the *Library of Congress Demographic Group Terms*, are introduced. The section on ontologies has also been significantly revised. Chapter 11 covers theory of categorization and how this translates into classification and taxonomies. It also contains brief overviews of types of classifications and the major classification systems used in the United States.

One method of determining aboutness continues to be covered in Appendix A: An Approach to Subject Analysis; this appendix has been updated. Because arrangement and display of information are so important to users' retrieval of information, two appendices are devoted to these issues: Appendix B: Arrangement of Physical Information Resources in Libraries, and Appendix C: Arrangement of Metadata Displays. Appendices D and E provide encoded metadata examples that were too large to be placed within the appropriate chapters. And, of course, the glossary and bibliography have been significantly updated to reflect all of the aforementioned updates.

NOTES

1. Arlene G. Taylor, *The Organization of Information* (Englewood, CO: Libraries Unlimited, 1999), xvii.

2. Daniel N. Joudrey and Ryan McGinnis, "Graduate Education for Information Organization, Cataloging, and Metadata," *Cataloging & Classification Quarterly 52*, no 5. (2014): 525–26.

ACKNOWLEDGMENTS

The authors wish to acknowledge the assistance and support of a number of people who made our jobs easier. We thank Dr. Candy Schwartz for her invaluable feedback on the sections dealing with digital collections, the Semantic Web, and other online settings. We also wish to acknowledge the direct contributions of the following scholars, who reviewed, edited, and updated material for specialized sections in the book:

Dr. Naresh Agarwal	Knowledge management
Jesús Alonso Regalado, MLIS	Collection development
Dr. Kyong Eun Oh	Personal information management
Dr. Amber Stubbs	Natural language processing
Dr. Katherine M. Wisser	Archives, indexing, and records management

We are also immensely grateful to Dr. Blanche Woolls and Emma Bailey, in their respective roles at Libraries Unlimited, for running interference for us when we were confounded by unexpected publishing perils. The authors are tremendously appreciative for the assistance and support provided by the people listed above. Each author, however, also has some personal acknowledgments to share.

Daniel N. Joudrey: I would like to thank the many people who have played a role in the development and writing of this work. First and foremost, I am grateful for the presence of Dr. Arlene G. Taylor in my life, without whom I would not be the person or the professional that I am today. She has been my teacher, my mentor, my friend, and my writing partner. She has been generous, supportive, encouraging, challenging, and thought provoking since the moment I met her in the fall of 1998. She has been an inspiration in my career as a teacher, a scholar, faculty member, and a conscientious, thoughtful author. It has been an honor and a pleasure to work with her over the years. Thank you for entrusting "your baby" to me.

I also must thank people associated with my home in academia—the School of Library and Information Science (SLIS) at Simmons College, Boston, Massachusetts. I thank my dean, Eileen Abels, and my brilliant and dedicated colleagues at SLIS, particularly Dr. Rong Tang, Dr. Melanie A. Kimball, Dr. Jeannette Bastian, Dr. Michéle V. Cloonan, Jeanne Wallace-Buckley, and, of course, my colleagues mentioned above (Kathy, Kyong Eun, Amber, and Naresh). I would particularly like to say something to the soon-to-be-retiring Dr. Candy Schwartz: "You have been my rock at Simmons. Thank you for everything. I will miss you so much when you retire. Please don't go!!" I also thank my students, particularly those who were in *LIS 415: Information Organization* during the Fall 2015 semester, for providing feedback on the previous edition. Their thoughts and input, and their ferocious desire for a more visual approach to the text, inspired changes found in this edition.

On a more personal note, I would like to say the writing of this book has not been easy. During the writing period, among my family and closest circle of friends, there were 17 hospitalizations and three devastating losses. Much of my energy had to be focus on my loved ones, and that delayed this book. At times, I did not think I would be able to complete the task, but with the support of my partner, family, and friends, I was able to focus on the work. I thank my mother Linda, my brother Matt, and my sister Christina for their love and their care, even as they were grieving themselves. They sustained me during these highly stressful and emotional times. This is also true of my closest friends, Tracy Ortman, Linnea Johnson, and Marcy Slack; I don't know what I would have done without them. I would especially like to thank my partner, my friend, my one true love, Jesús Alonso Regalado, Subject Librarian for Romance Languages and Literatures, and Latin American, Caribbean, and U.S. Latino Studies at the University at Albany, State University of New York. I have so much to thank him for, including absolutely *everything*. During these difficult times, he often put my needs above his so that I could focus on the book. I thank him for inspiring me with his expertise, his dedication, and his complete and total love of librarianship. I thank him for his impromptu reference services (he is *still* the best librarian that I have ever known). And I thank the universe for Orientation Day at the School of Information Sciences at the University of Pittsburgh in August 1999. It started my career and it is the day that I met Jesús—the two best parts of my life.

Arlene G. Taylor: I am so very grateful for the assistance of all who are already named, but among those, I am personally grateful for the friendship of Dr. Blanche Woolls. She has understood and helped with my personal challenges of the last couple of years beyond what is expected of an editor; but, then, we have known and supported each other for 25 years through a number of ups and downs!

I would like to thank Dr. Daniel N. Joudrey (Danny) for taking the lead on this edition of the work ("my baby"). I consider the work to be one of my most important contributions to the field, because I wrote the first edition from my course notes and received an award for it. I had been teaching "beginning cataloging" to students, many of whom I realized were not going to be catalogers. Those students needed to understand the underlying concepts of how information is organized, in all the ways that exist alongside catalogs, and so I reorganized my beginning classes to reflect this and then made the decision to write my notes into a book. For the book published in 1999, I was awarded the 2000 ALA/Highsmith Library Literature Award for, among other things, "providing a solid and impressive introduction to the theory, principles, standards, and tools behind the organization of information." Choosing a successor for this legacy was an important task. Danny began to assist with updating information for the second edition published in 2004. He served as co-author for the third edition published in 2009. And now he has accepted responsibility as the lead author for the fourth edition. I am immensely grateful for his lead in updating this edition. I have also enjoyed our friendship.

I again thank my husband, Dr. A. Wayne Benson, who has stayed by me for more than 30 years while I was engrossed in numerous books and articles. And to my friends in our retirement community who keep asking me "Is it done yet?" I say thank you for remembering and caring!

CHAPTER 1

ORGANIZATION OF RECORDED INFORMATION

Merriam-Webster's dictionary gives several senses in its definition of the word *organize.* The one that we are interested in first is "to form into a coherent unity or functioning whole: integrate." And secondarily, we use *organize* in the sense of "to arrange or order things so that they can be found or used easily and quickly: to put things into a particular arrangement or order" and "to arrange elements into a whole of interdependent parts."[1] This book, then, is dedicated to explaining the processes of forming unity and arranging or ordering elements for the purpose of retrieval. It examines the activities carried out and tools used by people who work in places that accumulate *information resources* (e.g., books, maps, documents, datasets, images) for the use of humankind, both immediately and for posterity. It discusses the processes that are in place to make resources findable, whether someone is searching for a single known item or is browsing through hundreds of resources just hoping to discover something useful. Information organization supports a myriad of information-seeking scenarios.

This chapter provides an overview of the field of the organization of recorded information. Terms used here that might not be readily familiar to those who are new to the field of *library and information science* (LIS) or its specialization *information organization* will be explained in more detail in later chapters. In the meantime the reader may find definitions of most unfamiliar terms in the glossary of this book.

THE NEED TO ORGANIZE

There seems to be a basic human drive to organize. Psychologists tell us that babies' brains organize images into categories such as *faces* or *foods.* Small children do a lot of organizing and matching during play. For example, children may divide their toys by shape, color, or type before playing or as a form of play in its own right. Many toys for toddlers are sorting and stacking toys that are designed to help young minds develop the ability to categorize—an important part of human cognitive functioning. As we grow, humans develop more sophisticated cognitive abilities to categorize, to recognize patterns, to sort, to relate, and to create groups of things and ideas. Cognitive scientist Steven Harnad has even said, "Cognition is categorization."[2]

With some individuals the need to organize is much stronger than with others. Those who operate on the maxim "A place for everything and everything in its place" may not be able to begin to work until the work surface is cleared and stray objects have been put in place. That is, such a person may have to be organized before beginning a new project. But even those whose workspaces appear to be cluttered or chaotic may have some organization in their heads. Such persons usually have some idea, or perhaps certain knowledge, of what is in the various piles or collections of stuff. In terms of personal information management, both popular culture and some scholarly research categorize people as *filers* or *pilers*, depending on how they organize their own materials.[3] (The two categories are rather self-explanatory!) Regardless of one's personal style, however, human learning is based upon the ability to analyze, organize, and retrieve data, information, and knowledge; to recognize patterns; to compare experiences, concepts, and ideas; and to process the relationships among all of these.

Why do we organize? In addition to the sheer pleasure of having everything in its place (whether that is in a folder, on a shelf, or in a particular pile), we organize for more significant reasons. We organize to

- **Understand**. Organization helps us to make sense of many things in daily life. For example, one

- **Understand**. Organization helps us to make sense of many things in daily life. For example, one can live in Cambridgeport, a neighborhood in Cambridge, a city in Middlesex County, which is a political division in the Commonwealth of Massachusetts, which is a state located in New England, the northeastern-most part of the United States. There are formal and informal organization systems of geography that help us to understand where we are and to communicate geographic locations to others.
- **Save time**. We organize to be quick and efficient. We want to access our information, things, services, and the like, in a timely manner. If, for example, the documents in an archival storage box are not grouped according to categories or series, we may have to sort through hundreds or even thousands of pieces of paper to find what is needed. That would take too much time.
- **Collocate**. We organize to bring similar things or ideas together into groups. For example, if we have many books on the same subject or sound recordings in the same musical genre, it is helpful to label them consistently so that they can be searched and presented together. When we find a book that is useful or some music that is appealing, we can discover similar resources nearby.
- **Retrieve**. Most of all, we organize because we need to retrieve. The world is filled with objects and ideas. We need to find them, whatever and wherever they are. Organizing makes this possible.

Information organization is employed in formal information environments such as libraries and archives, but many of our everyday mundane actions are dependent on the rest of the world around us being organized. Kitchens are organized so that cooking equipment is easily accessible and foodstuffs and spices can be used as needed. (Who does not want an organized spice rack?!) Items in a newspaper are organized into specific sections so that similar stories are presented together and the comics are separated from the obituaries. Personal music collections are organized so that we can find just the right song when we want it. Workplaces are organized so that documents are retrievable and work can be done. Learning processes are organized so that relationships among ideas can be used to assist the learner in recalling the learned material. Auto body shops are organized so that mechanics can find just the right part to keep the wheels from falling off a car while it is being driven down the highway. Organization is a very important part of everyday life; it can be key to efficiency, progress, and even survival!

Efficient and effective retrieval of information is dependent upon its having been organized. Information is needed in all aspects of life—for example, for health reasons, for entertainment, to understand each other, to learn about one's relationships, to fix things that are broken, or simply to expand our knowledge. Some of this information has already been assimilated and is in one's knowledge reserve, while other information has to be sought. If this information is not organized, it is difficult, and at times impossible, to find. (Can you imagine the inconvenience of using a printed dictionary that is not arranged alphabetically?) So we have many tools that are organized to aid in the process of finding information that we need: directories, dictionaries, encyclopedias, bibliographies, indexes, catalogs, databases, and the like.

The reasons we organize information can be summarized by using the language found in *Functional Requirements for Bibliographic Records* (FRBR), a conceptual model from the International Federation of Library Associations and Institutions (IFLA) that identifies key components of the universe of recorded information. FRBR states that we organize our resource collections so that users can accomplish the following tasks:[4]

Task	Explanation	Example
Find	**to search for entities that match specific criteria**	**to find Jorge Luis Borges's *El Hacedor* in the catalog through a specific search or through browsing**
Identify	**to confirm that an entity corresponds to the one sought; to tell similar entities apart**	**to identify that the author found is the Argentine Jorge Borges who wrote *El Hacedor* and not the Brazilian Jorge Borges, a biologist who writes about sustainability**
Select	**to choose a resource that is appropriate to the user's needs**	**to select the version of *El Hacedor* that is in a language understood by the reader**
Obtain	**to gain access to the resource described**	**to obtain *El Hacedor* from the library stacks or to request it from another institution**

Recently, in an update to the FRBR model (known as the *IFLA Library Reference Model*), another user task has been added:[5]

Explore	**to discover resources and entities and to gain greater understanding of**	**to explore other works created by Borges, which leads to the discovery of *Ficciones* and *Elogio de la Sombra***

resources

Organization of information also allows us to save for posterity copies of all kinds of works that result from human endeavors (e.g., books, art works, sound recordings, films, correspondence, tweets, government documents). Libraries, archives, museums, and other cultural institutions have been doing this for many years and have now moved into the online world. Although some comparisons occasionally may come from the business sector, generally this book is not concerned with the organization of information, resources, or materials in commercial enterprises and other profit-driven ventures that may have put together collections of items for the purpose of sale, rather than collecting items in order to benefit humanity and our collective knowledge reserve.

THE NATURE OF INFORMATION

Consider the following terms: *understanding, data, knowledge, wisdom,* and *information.* If you were asked to place these terms in order, how would you do it? Would you rank them from the lowest level of thinking to the highest? In alphabetical order? By the length of each word? In some other way? Clifford Stoll, in his book *Silicon Snake Oil: Second Thoughts on the Information Highway,* discussed these words that we use to indicate different levels of comprehension of symbols.[6] His order, indicating symbols from the least meaningful to the most meaningful, is

Data
Information
Knowledge
Understanding
Wisdom

(His order just happens to be in alphabetical order, too!)

Which of these are we organizing in libraries, museums, archives, and other such institutions? At various times, there have been spirited discussions over what it is that we are organizing. Some believe we are organizing information, and others believe we are organizing knowledge. The authors' bias is evident from the title of this book. It seems to us that we can use our knowledge to write a book, but until you read that book, understand it, and integrate it into your own knowledge, it is just information. That is why we believe we organize information—so that others can find it, read or otherwise absorb it, and use it to add to their own reserve of knowledge. Notice that in the preceding sentences, we said that you read, understand, and then integrate the information into your own knowledge. So there is some question about Stoll's putting understanding *after* knowledge. Perhaps the two concepts are intertwined. You need to have some understanding in order to incorporate something into your knowledge, but you must have a certain amount of knowledge in order to understand new things.

According to several dictionaries, *knowledge* exists in the mind of an individual who has studied a subject, understands it, and perhaps has added to it through research or other means. The same dictionaries indicate that *information* is the communication or reception of knowledge. Such communication occurs in great part through the recording of the knowledge in some fashion. People write, speak, compose, paint, sculpt, and attempt to communicate their knowledge to others in many other ways. This book, for example, is a representation of the authors' knowledge, but it is not a complete representation of our knowledge of the subject. It is, no doubt, an imperfect representation, in the sense that some concepts may not be explained as thoroughly as we truly understand them. However, it is not a representation of the reader's knowledge until the reader has read it, processed it, and understood it. That is, it is information that can be placed into a scheme of organization from which it can be retrieved for study by those interested in increasing their knowledge of the subject. Thus, we have chosen to use the term *information* rather than *knowledge* as our expression of what we believe we organize when organizing for the benefit of others. This is not a rejection of *organization of knowledge*, however. The knowledge existing in the brains of people is being harnessed in many situations. Authors work on organizing their own knowledge every time they write. The knowledge of reference librarians is used in an organized way when they assist patrons in answering questions. The phrase *knowledge management*, discussed below, has come into use in the administration of organizations. And, in recent years, some information professionals have begun to refer to classification schemes, thesauri, and other controlled vocabularies as *knowledge organization systems* (KOS).

The Nature of the Organization of Recorded Information

As mentioned earlier, this book addresses the organization of recorded information, as other means are necessary to organize information that has only been spoken, heard, or thought about. Recorded information, however, includes much more than just text. Video and audio recordings, images, notated music, and cartographic representations are all examples of recorded information that go beyond text. Therefore, instead of using words such as *book* or *item* to refer to describable units of information, the terms *resource* and *information resource* are used interchangeably in this book. A resource, thus, is an instance of recorded information; it can be anything from a 140-character tweet to a collection of 140 rock and mineral samples.

Despite the various forms in which information may be recorded, all information resources have some basic attributes or properties in common, such as *title* (what we call a resource), *creator* (who is responsible for it), and *subject* (what it is about). These attributes (and others) are recorded to help organize information; collectively, these attributes can be referred to as *metadata.* Metadata, in its most informal but most prevalent definition, is “data about data.” This means that the attributes used to describe an information resource are metadata about that resource. Metadata is discussed throughout this book, particularly in Chapter 5, and specific aspects of metadata are addressed in Chapters 7–11.

Ronald Hagler, in his book *The Bibliographic Record and Information Technology,* identified six functions of bibliographic control.[7] *Bibliographic control* (more often referred to as *information organization* today) is the process of describing information resources and providing name, title, and subject access to the descriptions, resulting in records or individual metadata statements that serve as surrogates for the actual items of recorded information. These statements or surrogate records (sometimes called *bibliographic records* or simply *metadata*) are usually placed into information retrieval tools, where they act as pointers to the actual information resources. The descriptions provide users with enough information to determine the potential value of the resources without actually having to view the items directly. Metadata is stored in a variety of *retrieval tools* (a collective term used to describe bibliographies, catalogs, indexes, finding aids, museum registers, bibliographic databases, search engines, and other tools that help users find information resources). Hagler’s listing reflects the purpose of his book—that is, the emphasis is upon the work of librarians, particularly *catalogers* (i.e., librarians responsible for managing and creating records for the library catalog). However, the list, presented and enlarged below, with wording altered to be inclusive of all types of recorded information, reflects the major activities involved in organizing recorded information.

1. *Identifying the existence of all types of information resources as they are made available.*

A book may be published, a streaming video released, or a website may be established, but if no one knows of its existence except the parties involved in its creation, that resource will be of no informational use to anyone. Existence and identity can be made known in many ways: publishers’ announcements (through various social media platforms, mailers, or email), reviews (both formal and informal), and subject-related listings, to name a few. Most publishers create sales catalogs listing their products with accompanying marketing blurbs. Some journals send regular announcements, outlining their contents, to let readers know when a new issue is available. Some organizations, publications, and content creators allow people to sign up to receive notifications (through email, RSS feeds, or other mechanisms) about new resources, information available at their websites, and so on. Reference tools such as *Books in Print* are products of this first function of bibliographic control.

2. *Identifying the works contained within those resources.*

In many cases, one information resource is equal to one intellectual, literary, or artistic work (e.g., a DVD that contains one feature film). However, in some cases, a single resource may contain numerous individual works. For example, a DVD may be a compilation of a dozen short films or an online portfolio may contain a collection of several hundred digital photographs. Both the compilation and the portfolio are considered to be works, but the individual short films and artistic images contained within are works in their own right as well. The level at which resources are described depends upon how much *granularity* (i.e., how much detail) is desired. As another example, an online resource about a famous person may have individual digitized works by the person, biographical material, a bibliography of the person’s works, a curriculum vitae, accounts of the life or works of the person written by contemporaries, accounts of historical events contemporary to the person’s life-span, and other such components. The writings about the person and those about historical events may be important works in their own right and may need to be identified separately.

3. *Systematically pulling together these resources into collections.*

Traditionally, the activity of creating collections has been thought of as the province of institutions such as libraries, archives, and museums. But collections have always been created in many other situations. For example, personal collections can be created because of an intense interest in a particular kind of information, office collections of internal documents and information are needed to carry out the work of the office, university departments may collect materials needed for teaching in a particular discipline, and so forth. Whereas in the past these collections might have remained hidden from view, it is now easy to make these collections known publicly through electronic means.

Also, collections today often include digital resources that are not held or owned locally. Many institutions procure the right to allow their users to access resources online. Some such resources are accessible *only* electronically, while others are also available in print. Part of the collection development process is determining whether such resources need to be added to one's collection in a more permanent way. Most institutions wish to collect in a purposeful way. Unless a collection is well organized, it can be difficult to determine which new works will enhance the collection. Developers of collections rely on metadata to avoid adding something that is already owned.

4. *Producing lists of these information resources prepared according to standard rules for citation.*

Lists created in the process of describing information resources include, among others, bibliographies, indexes, library catalogs, archival finding aids, and museum registers. These tools are important for the retrieval of individual information resources, because if one is looking for a known resource—especially a tangible one that has a physical location—it is necessary to find it listed somewhere. Such lists may still be in print form, but over the past few decades the preference has moved gradually toward providing online access to them, often exclusively. These lists have varying levels of complexity. Some, such as bibliographies, may be quite simple and fairly easy to use; others, such as library catalogs, may contain more complete and more complex information that requires greater sophistication in methods of access.

The way that each resource is described in a particular list is determined by the set of guidelines used by the institution creating the list. To ensure consistency and accuracy, resources are described following the instructions in a *content standard* (i.e., a rulebook or instruction manual for metadata creation). Different types of institutions follow different content standards. In the United States, for example, libraries typically use *RDA: Resource Description & Access*, archives use *Describing Archives: A Content Standard* (DACS), museums may describe their collections using *Cataloging Cultural Objects* (CCO), and the creators of bibliographies may use any of several available style manuals (e.g., *Chicago Manual of Style*, *MLA Handbook*). Most content standards, at a minimum, provide an overview of the description process, an indication of what data elements are mandatory and/or desirable, and some guidelines for how to record the metadata. Not all metadata creators, however, adhere to a formal content standard when producing descriptions.

5. *Providing name, title, subject, and other useful access to these information resources.*

The successful retrieval of information resources through lists depends on the inclusion of sufficient metadata. The activity that adds the most value to the usefulness and retrieval potential of a collection is the provision of authority-controlled name, title, and subject access points to the descriptions of the information resources. An *access point* is a name, word, phrase, or identifier chosen by a cataloger or indexer and placed in a particular field in a record that describes a resource. It may then be used to obtain that record from a retrieval tool or other organized system. *Authority control* is most often practiced by using a unique character string to represent each name, work, or subject, in order to achieve consistency within the catalog or other retrieval tool. It also involves creating explicit relationships between different names, works, or subjects.

Keyword access, by contrast, can be provided more or less automatically and "on-the-fly"—that is, any information in digital form can be found by searching for a word that appears in the resource. However, keyword searching is not always efficient or precise, and not everything on that topic may be retrieved, due to a lack of semantic controls. As the size of the collection being searched increases, results of keyword searches become less and less satisfactory because of *infoglut* (i.e., information overload).

More satisfactory retrieval comes from being able to search for specific names, titles, and subjects that have been chosen and constructed under authority control to bring together variant forms and related terms. If a person has been identified by different names and/or forms of name (e.g., Jacqueline Bouvier, Jackie Kennedy, Jacqueline Kennedy Onassis, and Jackie O.), and if that person's name is brought under authority control, then one name or form of name is chosen as the preferred one and the others are documented as cross-references to it. Thus, a search for one form of the name will retrieve information resources related to the person regardless of which name or form of name appears in a particular resource. For example, the user can search either *Kennedy, Jackie* or *Onassis, Jacqueline Kennedy*, and still receive relevant search results. This may be accomplished through the use of a direct reference, such as

You searched: *Kennedy, Jackie*

In this catalog, please search: **Onassis, Jacqueline Kennedy, 1929–1994** instead.

Or the retrieval tool may be configured to search for a preferred name automatically and seamlessly when one of the references is used in the user's search. Additionally, if a work has been given different titles in various manifestations, then the most common form of title is usually selected as the access point for the work. All the variant titles are connected, and a search for one of the titles should allow users to navigate to every version of the work, no matter the title given to an individual edition or version of the resource. In other words, a search for *Hamlet* should also retrieve results for *Amleto, Principe di Danimarca*; *Hamlit*; *Gamlet*; and *The Tragicall Historie of Hamlet, Prince of Denmark*, if all those variants are found in the list being searched. And, if a retrieval system uses controlled vocabulary to represent subject concepts, a search for a word with more than one meaning (which is a characteristic of most English words) often will allow differentiation among the various meanings (e.g., **Bridges (Computer networks)** versus **Bridges (Dentistry)**) and will direct a searcher to broader, narrower, and related terms. It will also bring together, under one term (e.g., **Periodicals**), all the synonymous and nearly synonymous terms that may be used to express a concept (e.g., journals, magazines, periodical publications).

Authority-controlled access is of little use, however, unless systems are designed to take advantage of it. The retrieval systems designed for searching and displaying organized information are not always sophisticated, intuitive, or easy to navigate. Without the support of thoughtful user-oriented system design, the benefits of information organization can be diminished or lost entirely.

6. *Providing the means of locating a resource.*

Location of information resources has been, for more than a century, a value added by institutions with collections. The printed catalogs or other lists created in these institutions gave information on the physical location of the resource. Most library catalogs, some museum information, and many archival finding aids are now available online. In many library online catalogs, circulation information is available so that if a resource has been taken out of the library, that information is made available to the patron. This does not apply to archives and museums because they generally do not circulate their collections to the public. (Sorry, but you cannot take Dali's *Persistence of Memory* home for the weekend from the Museum of Modern Art!) Bibliographic networks (e.g., OCLC Online Computer Library Center) are also helpful because they indicate which institutions own particular resources. One can discover which libraries own the resource sought, and then learn from individual library sites whether the resource is available or is on loan to a patron (and sometimes whether the resource is on order and when it is expected to arrive).

Traditionally, bibliographies and indexes have not given location information. Bibliographies list information resources that exist somewhere but seldom tell where you may find copies of them. Indexes give the larger resource in which a smaller work being listed can be found (e.g., the journal in which an article can be found), but they do not give the physical location of the larger resource. All of this is still true for tangible resources, but for online resources it is now common to give the direct location in the form of a *Uniform Resource Locator* (URL)—a web address for the resource.

In order to accomplish Hagler's six functions of bibliographic control, adequate and sufficient metadata is needed. Without this metadata, information resources could not be made known to the world, smaller works could not be identified, collections could not be created, consistent lists could not be produced, access could not be granted, and locating the resources would be impossible. Metadata is necessary in order to meet these and other functions as well. Metadata, however, will not look the same in all cases. There are differences in how metadata is created for different types of materials, in different environments, among different communities, and in different contexts.

ORGANIZATION OF INFORMATION IN DIFFERENT CONTEXTS

There are many contexts in which there is a desire to organize information so that it will be retrievable for various purposes and so that at least some of it will be preserved for posterity. The ones to be discussed here are

- Libraries (of all types)
- Archives
- Museums
- Online settings and contexts, such as
  - The Internet
  - The Semantic Web
  - Digital collections
  - Information architecture
- Indexing and abstracting
- Records management
- Personal information management
- Knowledge management

Libraries

We consider libraries first because they have the longest tradition of organizing information for the purpose of retrieval and for posterity. In most libraries, information organization activities are centered in the technical services department. The term *technical services* (often shortened to *tech services* in casual conversation) refers to the "behind-the-scenes" activities that address developing, acquiring, organizing, and preparing a library's collection of resources.[8] Additionally, there may be other units within the department focused on serials, digital resources, preservation and conservation, archives, or technology. There is not universal agreement, however, on just which units belong in the technical services department; each library may have a different configuration of units placed under the tech services banner. For example, in some libraries, collection development may be more closely associated with reference and *public services* (a library's front-of-house operations) or it may be its own department. In Figure 1.1, a very basic outline of selected technical services activities is provided. The processes outlined in the figure are addressed further in the text.

As mentioned earlier, the process of organizing recorded information begins with collections. Collections in libraries are created through the collection development process, which may utilize any of the following methods:

- **Selection**: the process in which collection development librarians (also known as *subject librarians* or *bibliographers*) learn about the existence of works through vendors' product catalogs, reviews, publishers' announcements and websites, requests from users of the library, and the like, and then choose the most appropriate materials for the collection;
- **Gifts and Exchanges**: *gifts* are information resources that are donated to a library from an interested party (e.g., from a trustee, a local organization, or a retired faculty member); *exchanges* refers to when a library has a mutually beneficial trade agreement with one or more other libraries to swap its duplicate or unwanted resources for more desired resources;
- **Approval Plans**: a method in which a library contracts with one or more vendors to receive new resources according to pre-selected profiles outlining the collection's needs; and
- **Patron-driven Acquisitions**: also known as *demand-driven acquisitions*; the process in which a library obtains certain resources (usually electronic) only after the need for them has been definitively established (e.g., a library may provide access to a large number of e-book records in the catalog, but the library only purchases an individual e-book after it has been accessed by a patron).

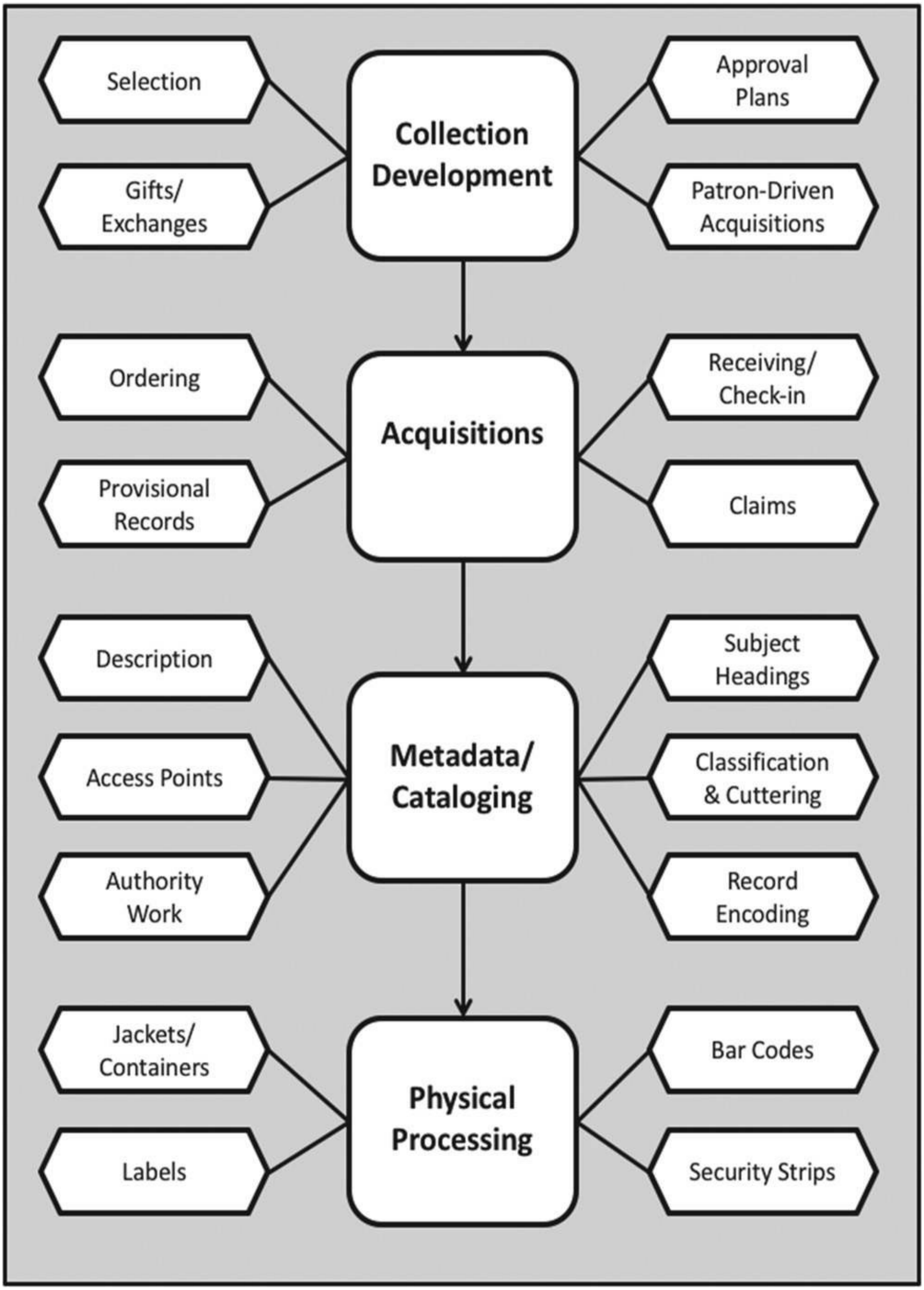


Figure 1.1. An Overly Simplistic Library Technical Services Workflow Diagram.

Less frequently, but still practiced in some academic and research libraries, collection development librarians may take resource-buying trips to purchase special materials, particularly difficult-to-find foreign language materials.

Acquisitions is another function related to the development and organization of the collections. The acquisitions unit, among other things, is responsible for managing orders and budgets (e.g., ordering materials, paying the bills, overseeing standing orders, monitoring budgets). Additionally, they may be responsible for receiving the resources, checking them in, claiming missing items, and managing returns. Acquisitions librarians may also be responsible for obtaining or creating temporary catalog records at the time of ordering materials.[9] After the materials have been checked in, these provisional records are expanded upon in the next process.

Once resources arrive at the library, they must be organized and integrated into the collection. This process is called *cataloging*, and its goal is to create a multifaceted list of resources to which the library can provide access. Cataloging comprises two major activities: *descriptive cataloging* and *subject cataloging*. "Descriptive cataloging describes the makeup of an information resource and identifies those entities responsible for its intellectual and/or artistic contents without reference to its classification by subject or to the assignment of subject headings, both of which are the province of subject cataloging."[10] Descriptive cataloging involves the following activities:

- **Creating a Description**—This process entails recording the most important attributes of a resource to allow users to find, identify, select, obtain, and explore it. The description, often, is a straightforward representation of the "facts" of the resource addressing *who*, *what*, *where*, and *when* (though perhaps not *why*). A minimal description includes the following attributes (if applicable): title of the resource; statements of responsibility; version or edition information; dissemination information; the type of resource; its extent, dimensions, and size; and standard numbers associated with the resource.
- **Choosing Access Points**—This activity ensures that the names of the most responsible parties, as well as various titles associated with the resource, are enumerated in the metadata, and these access points can be used to find the resource in the catalog. Access points may include
  - names of creators (e.g., authors, artists, national or local government bodies, compilers, composers, families, musical groups);
  - names of contributors (e.g., editors, translators, illustrators, abridgers, performers, sound designers);
  - titles given to the resource; and
  - titles of related resources (e.g., prequels, adaptations, series titles).
- **Ensuring Authority Control**—This activity provides consistency in how names and some titles appear as *authorized access points* (i.e., standardized character strings chosen to represent names or titles). For example, this means that catalogers consistently use the authorized access point **Onassis, Jacqueline Kennedy, 1929–1994** in the metadata for resources by or about Jackie Kennedy.

The description, choice of name and title access points, and authority control for names and titles are accomplished in most libraries in the United States and in many other countries through reference to *RDA: Resource Description & Access*.[11] This content standard is the replacement for the *Anglo-American Cataloguing Rules, Second Edition, 2002 Revision* (AACR2),[12] the previous content standard. RDA is based on AACR2 and on the conceptual models for bibliographic and authority data that have been developed by IFLA: FRBR[13] and *Functional Requirements for Authority Data* (FRAD).[14]

Subject cataloging, the second major component of library cataloging, entails two key activities: *conceptual analysis* and *translation*. Conceptual analysis is the process of determining the *aboutness* (i.e., subject matter) of a resource. Catalogers analyze a resource to determine the concepts, ideas, and entities discussed within it, as well as to determine its form or genre. A resource cannot be accurately described if it is not thoughtfully examined and understood. Translation is the process of transforming the resource's aboutness into controlled subject languages. It includes the following activities:

- **Choosing Controlled Vocabulary Terms**—Catalogers assign authority-controlled terms from a list to represent the aboutness of a resource as part of its metadata. Controlled vocabulary ensures consistency in subject representation and allows for *collocation* (i.e., bringing together like materials). As an example, a resource about dental bridges and crowns may be assigned one of the following sets of subject headings—which set is applied depends on the list used by the library:

- **Bridges (Dentistry)** and **Crowns (Dentistry)** from *Library of Congress Subject Headings* (LCSH),[15] or
- **Denture, Partial, Fixed,** and **Crowns** from *Medical Subject Headings* (MeSH).[16]

- **Choosing Classification Notations—**This process involves representing the aboutness of a resource through standardized notation. Classification ensures that similar resources are collocated in the catalog and on the shelves. For example, in the United States the above resource may be assigned
  - **RK666** (which stands for *crown and bridge work* in prosthetic dentistry) from *Library of Congress Classification* (LCC),[17] or
  - **617.692** (*dentures, bridges, crowns*) from *Dewey Decimal Classification* (DDC).[18]

When materials are organized into collections, physical entities have to be arranged in some fashion. They may be placed on shelves in the order in which they arrive, or they may be placed in some more meaningful order. They could be arranged in alphabetical order, the way that fiction and biography sections are arranged in many public and school libraries. Most resources, however, are arranged by classification. For tangible resources, classification notations are used to create location devices referred to as *call numbers*. This is done to identify the physical resource and to provide a unique shelf address for it. A call number is usually established by adding a Cutter number to the classification notation assigned to the resource. A *Cutter number* is a sequence of alphanumeric characters that usually represents the primary creator's name, or lacking such a name, the title. There are different approaches to establishing call numbers depending on the classification scheme used. For example, *Bridges and Crowns: An Introduction*, a book by D. N. Allan published in 1986, might be assigned the following call numbers:

	LCC	DDC
Classification Number	RK666	617.692
Cutter Number	A45 represents names beginning with the letters *a-l-l* in the Library of Congress's cutter table	*A417* represents the name *Allan* in the Cutter-Sanborn table
Other Components	*1986* is the date of publication	*b* is the first letter of the first non-article word of the title
Complete Call Number	RK666 A45 1986	617.692 A417b

Most records created through the cataloging process are encoded with the MAchine-Readable Cataloging (MARC)[19] format so that they can be displayed in a catalog, which is usually part of a library's larger automation tool known as an *integrated library system* (ILS) or a *library services platform* (LSP).

The process of cataloging just described is often referred to as *original cataloging*. Fortunately, it is not necessary for every information resource in every library to be cataloged originally in that library. Because libraries frequently acquire copies of the same resources, their catalogers can share metadata by adapting a copy of the original cataloging record created by another library for their own catalogs, a process commonly called *copy cataloging*. This is related to the idea of *cooperative cataloging* —the working together by independent institutions to share network memberships or to create cataloging that can be used by others. There are various ways to obtain copies of existing cataloging records, such as through membership in a consortium or a bibliographic network (e.g., downloading records from OCLC), from vendors (e.g., paying for catalog records to be sent along with purchased resources), or from libraries that make their metadata freely available (e.g., downloading a catalog record created by the Library of Congress from their catalog).

Once resources have been chosen, received, and added into the catalog, the physical items are sent to the physical processing unit so that they can be prepared to be added to the collection. This involves removing or adding book jackets, placing security strips in or on materials, placing call number labels and barcodes on the resources, sending a resource to the conservation/preservation department if it is not in good shape, and so forth. Once the resources are processed, they may be shelved in the stacks to be accessed by users and circulated, if appropriate.

The two major outcomes of cataloging are (1) the arrangement of collections and (2) the creation and maintenance of the catalog that provides the primary means of access to the collections. The catalog is able to show what exists in the collection by certain creators, having certain titles, or on certain subjects. It also collocates all of the works of a creator and all the editions of a work, and all works on a subject, even though they might not be brought together in the collection.[20] Finally, the catalog provides some kind of location device to indicate where in the collection the item will be found, assuming it is not circulating to a user.

Before online catalogs existed, typically the library's main card, book, or microform catalog was supplemented by other catalogs. Catalogs for departmental libraries, serial record holdings, special formats catalogs, and shelflists containing location information for specific copies of a resource were the most common. Today, all of these have been incorporated into one database in most online catalogs. Because most online catalogs are part of integrated systems, circulation information can accompany each catalog record.

Until recently the online catalog primarily contained records only for resources physically held by the library system. As libraries have entered into cooperative relationships, the longstanding principle of the catalog showing "what the library has"[21] has eroded. In *union catalogs*—catalogs that contain records from more than one library—the concept was expanded to show what at least one of the cooperating libraries had. More recently, the addition of records for online resources has meant that catalogs now contain records for what the library provides access to (in addition to what the library has).

Library portals accessible from online catalogs provide users with a way to locate all the information content and services that they have the authority to access. The portal server presents an authentication screen to users prior to entering the system or when they are trying to access protected information resources. If the user name and password are accepted, the user can have access to whatever resources and services are allowed by the user's status. An academic institution, for example, may have licenses for its users to access many different online databases. Formerly, one had to learn the access protocol for each database and enter a different user ID and password. Through the controlled access of a portal, an authorized user may be able to search these databases by just clicking to enter them. A library portal may have access links for local resources, remote resources, reference help, and personal patron information, for example. Local resources may be divided into books, journals, databases, digital collections, and course reserves, and all of these may be accessible through searching the catalog. Reference help may include online reference tools, links to online search engines, and a virtual reference desk with either real-time or email access to a reference librarian. Personal patron information may include lists of materials borrowed, saved searches, and personalized alerts.

Online catalogs also can be gateways to outside resources. A major improvement to online catalogs was the addition of the ability to access remote Internet resources directly. Now, a URL in a catalog record can provide immediate access to additional metadata (e.g., tables of contents), to additional or complementary content (e.g., a supplement, an image of a book cover), or to the entire full-text resource that is represented in the description, if available (e.g., a link to a PDF version of a resource). Catalogs may also connect to the catalogs of other nearby institutions or even to bibliographic networks (e.g., OCLC) that can show where a resource may be found if it is not in the local catalog. The resource can then be requested through *interlibrary loan* (ILL) or some other form of document delivery.

In addition to searching the inventory of tangible library resources found in the catalog, bibliographic and full-text databases may be accessed through a catalog's *discovery layer* (also referred to as a *discovery interface*, *discovery tool*, or *discovery service*). A discovery layer is a technological add-on to the catalog. It allows users to have a greater variety of interactions with a library's information stores. For example, through the discovery layer, patrons may be able to search for books, journal articles, and digital objects all in one search, or they may be able to browse keyword search results that have been sorted into *facets* (i.e., fundamental aspects, features, or characteristics of resources) such as genre, resource type, creator, subjects, publication date, location, and so on.

An influence on the organization process in libraries is found in the reference process. Libraries are organized so that information can be retrieved. In the reference process the success of the organization is tested. If it is found to be difficult to use, some of the organization process must be redone. Administrative services in libraries also must be concerned with the organization of information. Administrators are responsible for technological decisions that directly affect the organization of the recorded information in that setting. Administrators' decisions also affect the future, in which chaos will result if the activities and processes of information organization are not supported.

Archives

Libraries became more and more standardized throughout the twentieth century, with many information resources in a library being duplicates of resources in another, but this is not the situation in archives. An archives[*] usually consists of unique items. Therefore, it once was thought that standardization was not applicable or necessary. Archives could not take advantage of previously created metadata (i.e., use metadata records created by other agencies) because they were not describing materials that were also owned elsewhere. More recently, however, archives have seen significant standardization in their descriptive practices.

Archives preserve records of enduring value that document organizational or personal activities accumulated in the course of daily life and work. Archives may contain *any type* of material, but some of the more common types found in organizational records collections include such things as annual reports, correspondence, personnel records, and the like. Personal records might consist of such things as correspondence, manuscripts, and personal papers, or perhaps a collection of memorabilia or a scrapbook. Even though materials in archives often are thought to be old, this is not necessarily so. Further, archival materials can be in many different formats: texts, graphic images, sound recordings, moving image recordings, on paper, or in analog or digital forms.

Despite the lack of formal organization systems, archival materials have been arranged and described for centuries. Unlike library materials, archival materials are arranged and described in groups. Until the last few decades, each archives chose its own way to organize the information, particularly regarding the level of control and depth of description. There have been several schools of thought as to the best approach to organizing the materials in archival collections and the creation of appropriate representations for those materials. Two fundamental principles serve as a foundation for current archival standards: *provenance* and *original order*. Provenance, or *respect des fonds*, refers to the individual, family, organization, or institution that is responsible for the creation, maintenance, or use of the materials. This principle dictates that records of different origins should be kept separate to preserve their context. Original order directs archivists to maintain the organization or sequence of materials established by the creator of those materials. This may or may not be evident in the collection. If the original order has been lost or the collection was never truly organized by the creator or collector, then an archivist may need to construct a logical order for the collection. In establishing a logical order, whether original order or one established by the archivist, subgroupings (*series*) among the materials may be identified. The description of the collection should reflect the arrangement of those groupings. Most archives keep the contents of individual collections within the archives as a whole in original order, and the collections are maintained according to provenance.

Standardization and cooperation came to the archival world beginning in the 1980s because of increased interest in research involving documents and archival collections housed all over the globe. In the 1990s and 2000s, interest also grew in making descriptions of archival collections accessible online and in including summary descriptions of archival collections in the same databases with library catalog records.

Descriptions of archival materials can take different forms and perform different types of functions. An *accession record* summarizes information about the source of the collection, gives the circumstances of its acquisition (which are more fully treated in the donor file), and briefly describes physical details and contents for a collection. A *finding aid* provides detailed notes on the historical and organizational context of the collection and continues by describing its content, providing an inventory outlining what is in each box. It may also contain subject headings, authority-controlled access points, and physical details such as the presence of brittle or fragile materials. Finding aids can also be referred to as *registers*, *inventories*, or *container lists*. A separate record, which describes the collection (or a part or parts of the collection), also may be created for inclusion in a library catalog.

In order to meet the responsibilities of taking care of records of enduring value while at the same time providing access to those materials, archival materials are generally housed in boxes kept in closed stacks, accessible only to staff. There is no public browsing of archival repositories, so the arrangement of the individual archival collections in the stacks does not need to be classified as is usually true in a library with open stacks. In addition, the varied nature of archival collections would require such a broad classification as to diminish the utility of that assignment.

In the 1980s, when the archival world became interested in placing records for archival collections into library catalogs, a MARC format for archival and manuscript control (AMC) was developed (MARC-AMC). Although MARC-AMC is no longer a separate format (having been integrated into the MARC bibliographic format), MARC can still be used to code archival catalog records. In the late 1990s, the Encoded Archival Description (EAD)[22] standard was released and has been implemented internationally to encode finding aids so that they can be manipulated and displayed online.

The organization of archival materials is a fundamental activity in the management and provision of access to collections. Archival description serves many purposes. It provides guidance to researchers on the utility of the material for their information needs. It also assists archivists in the management and preservation of their collections. When collections are well organized and documented, archival materials can fulfill a wide variety of uses and user needs.

Museums (Art and Object Collections)

Although libraries and archives may contain some visual materials (notably photographs, but also slides, art prints, and the like), the collections of museums, art galleries, and other institutions that collect artifacts and objects of material culture consist primarily of visual material in two- or three-dimensional form. The organization of these materials traditionally has been for internal use only. Recently, however, research needs, institutional mandates, and general public interest have led such institutions to think about organizing their collections in ways that libraries and archives have for many years. Metadata schemas, such as *Categories for the Description of Works of Art* (CDWA);[23] controlled vocabularies, such as the Library of Congress's *Thesaurus for Graphic Materials* (TGM),[24] the Getty vocabularies,[25] and *Iconclass*[26] (a subject classification system for describing the narrative and iconographic content of visual works); and content standards, like *Cataloging Cultural Objects* (CCO),[27] have emerged in the last few decades as a response to museums' needs to better organize and disseminate their collections' information.

Museum art works or artifacts are usually obtained through an institution's formal acquisitions process. As is done in archives, accession records are created, although the practice in natural history museums differs somewhat. In natural history museums, artifacts are acquired largely from fieldwork, and a preliminary field record is made; if it is decided to keep the objects in the collection, accession records are then created. In some cases, groups of similar objects are described as a single lot that is given a single accession number. The curation of individual objects, which may not happen for some time, results in departmental-level catalog records with their own numerical sequences. In museums other than the natural history type, items are registered after being accessioned. Registration is a process much like cataloging in libraries. The records created by a museum, historical society, or special collections registrar's department serve as a catalog in that they establish organizational control over the art works and artifacts.

In museums, the provenance of an object or a collection of objects is important information and is essential in determining the name and other elements describing that particular object. Provenance means slightly different things in museums and archives. Provenance in archives drives the principles of organization. Collections are organized according to their provenance. In art history, the provenance of a work is connected to the ownership and authenticity of the work. Both provenance and physical condition of the object must appear with all other information about the object in the catalog or registration record. An aspect of creating records for museum objects and art that is very different from creating records for bibliographic resources is that the objects are often imperfectly known at the time of accession or registration. There may be an accumulation of information over time, some of which may be conflicting or contradictory, which becomes part of the documentation.

Description of visual material can be more difficult than description of textual material. There is more reliance on the perceptions of the person doing the describing. Often there are no words associated with items at all; it is necessary for the describers of such items to use their own words. A single record may have many more fields than does the usual library catalog record. Some fields that are typically needed for art objects that are not used in libraries are: materials, technique(s), provenance, exhibition history, installation considerations, and appraised value. Even with additional fields, it is not possible to anticipate all of the uses a researcher might find in works of art or artifacts. A street scene from a century ago may be useful to historians, architects, urban planners, cultural historians, medical researchers, sociologists, students of photography, or others. Systems have been developed that start with queries that use the text of the description; then query results allow the searcher to browse surrogate images.

Subject analysis also presents challenges for visual materials—an image does not tell in words what it is about, nor does the title of a visual work always clearly convey what the work is about. Additionally, the line between description and subject analysis is harder to draw. One might describe a work of art as being a painting of a woman in a blue dress holding and looking at a baby—this is a description. But if one gives the subject of the work as "Mary and Jesus," one has crossed the line into interpretation (unless this is the title of the work given by the artist). And if one uses a description like "maternal love," one is definitely interpreting.

A perceived barrier to shared metadata about museum objects has been the fact that museums hold unique objects. This is perhaps less true of natural history collections than other museum and art collections; although each specimen of a bug or bird is unique, each represents a class of organisms that can be identified to the genus–species level and is kept not for its unique qualities but for its representational qualities. In the case of these natural history specimens, there may be copy-specific notes, but this does not preclude the idea of reusing metadata for cooperative access. However, as was true of libraries when cooperative cataloging was first introduced, museum curators in other types of museums fear a loss of individual control and a diminished level of detail. Until the advent of initiatives like Artstor[28] and OCLC's WorldCat,[29] museums had been reluctant to give up their local terminology and idiosyncratic ways of organizing their information in order to participate in union catalogs or federated retrieval tools. This is now changing. Chapter 7 describes some cooperative museum projects that are currently under way.

Besides their major collections, museums can also have archives, records management programs, and libraries. A museum library may contain published materials that document or relate to the museum collections. Museum archives document the institutional records of the museum as it accumulates, displays, and creates programming based on its collections. And because museums are institutions with responsibilities for proving their tax status (often nonprofit), among other things, they also have records management programs to ensure that appropriate and adequate documentation is kept over time.

As with archival materials, museum collections are accessible only to staff. Much of the collection is stored behind the scenes while only a portion of it is on display at any one time. Behind the scenes, the items are numbered in such a way that they can be retrieved as needed. Persons responsible for the exhibits make heavy use of the system of organizational control. These collections are increasingly also being used for research by persons with diverse research needs, but mediated access has not required a change in strategy for organizational control.

Online Settings

In the 21st century, the Internet has become omnipresent in our personal and professional lives. The Internet—particularly the web—is a tool that we use for nearly everything. We use it to conduct business transactions, to manage our calendars and bank accounts, to find answers to quick questions, and to further investigate almost anything that we are curious about. The Internet, however, is not a panacea for all of life's information problems. The Internet is unpredictable and inconstant; it can be unkind and untrue; it is full of opinions that may masquerade as facts; and it is very messy. The Internet has been likened to a library where all the books have been dumped on the floor and there is no catalog.[30] For many years, efforts have been made to find ways to gain some control over it. One cannot say, however, that the Internet is organized, and this is unlikely to change immediately. Portions of it—certain types of online spaces and many individual sites—may be organized in some fashion, but they are not representative of the entire Internet. The next four subsections discuss some of these online settings: the general Internet, the Semantic Web, digital collections, and information architecture.

The Internet

In the mid-1990s, some librarians and LIS researchers enthusiastically and somewhat naively attempted to design projects to help organize the Internet—the *whole* Internet.[31] These efforts began in earnest only after the World Wide Web (or *the* web) became the face of the Internet. These professionals, however, were quick to realize that there was so much change so fast that many efforts were out of date in just a short time. Early on, libraries attempted to use traditional means of organization, but it became clear rather quickly that full-scale library cataloging was not particularly well suited for describing what often turned out to be ephemeral web-based resources.

Instead of creating individual records for each website, some librarians began to compile bibliographies of sites, some of which eventually became more formal directories or gateways to the Internet. An *Internet directory* is an organized collection of links to websites on particular topics. In many directories, human indexers curated the links at first, but as technology developed, some began to collect links through automatic means. INFOMINE, established in 1994, is an example of an Internet directory created by cooperating librarians from several academic institutions. It contained pages dedicated to various topics (e.g., cultural diversity resources, government information, maps and geographic data), each of which contained lists of links to recommended websites. The links could be browsed in alphabetical lists or searched using title words, creator names, subjects, or keywords. Its organization was similar to, but simpler than, that of the more widely known Yahoo! directory, which also debuted in 1994. Yahoo! favored a more hierarchical, drill-down method of browsing, where one started with one of the 14 top-level categories that were divided into myriad subtopics or facets. In any given category, one could easily find five or six levels of subtopics and, in some cases, many more. Hierarchically structured directories of websites, however, eventually lost favor to general search engines, such as Alta Vista and Google, after many improvements increased the efficacy of search engines. After two decades of operation, both Yahoo! and INFOMINE closed their directories in December 2014 after experiencing significant declines in usage in later years.[32] An example of an enduring Internet directory, at least at the time of this writing, is *SciCentral*, which provides up-to-date links to science-related news stories, journals, job resources, conferences, and so on.[33] Although a small number of gateways still exist, hierarchical directories are mostly relics of the Internet's past. As information architects Louis Rosenfeld, Peter Morville, and Jorge Arango state, "The findability techniques that were effective in the late 1990s (e.g., Yahoo!'s curated hierarchical directory) are ineffective today."[34]

Another attempt to control the information on the Internet resulted in the development of one of the world's most widely known metadata schemas. Librarians were part of the team of people who developed a metadata standard for describing online resources called the *Dublin Core* (DC).[35] It was designed as a relatively small set of basic metadata elements, because most websites and other document-like objects on the Internet did not require as much detail as was typically found in the library catalog. In the late 1990s, OCLC established a way for libraries to catalog online resources cooperatively and to have ready access to a database of metadata describing important web resources. This system, originally named CORC (Cooperative Online Resource Catalog), later evolved into OCLC's web-based interface for cataloging—Connexion. Today, in Connexion, resource description can be done either in the traditional MARC format or in DC.

Much work on organizing and searching the Internet has been done by persons other than librarians. For example, computer scientists and programmers have primarily developed search engines—the principal tool used to access web resources. Most people have a great appreciation for search engines, even though they may be somewhat frustrated that search engines are not more selective and precise. Most programs or agents (e.g., crawlers, robots, spiders) sent out to find sites to add to the indexes of search engines are able to index text only; videos, sound files, and images can be recognized as such, but they may not be indexed unless they also have accompanying text or labels. In addition to that limitation, these programs cannot analyze a site's purpose, history, policies, and so forth. In order to improve the situation, work on various kinds of metadata for digital resources is ongoing and important. Appropriate information could be gleaned by robots from metadata that has been added to a site by its creator or by someone trained in describing and analyzing resources; in the past two decades, however, misuse of metadata (e.g., addition of keywords that are popular words but have nothing to do with the content of the site) has kept many search engines from making extensive use of it. This misuse can be seen as unethical, homegrown attempts at *search engine optimization* (SEO)—an activity aimed at raising the visibility of websites (i.e., getting sites to be ranked higher in search engine results). Legitimate SEO activities focus on sets of best practices in marketing, web design, and content organization to achieve their goals, rather than on deception.

In recent years, the attitude toward including metadata in web resources has changed. Some of the world's most popular search engines (i.e., Google, Bing, Yahoo!, and Yandex) have started to embrace certain semantic technologies that embed small amounts of metadata into Hypertext Markup Language (HTML) code to make the data found within websites more meaningful to search engines and web crawlers. (See discussion of microdata below.) Properly used, metadata can include, among other things, information about creators, titles, subjects, and other traditional aspects of description; nontextual parts of a site; textual portions of the site; entities, roles, and relationships; and the site's purpose and history.

That initial desire among librarians and other information professionals to describe each piece of the *whole* of the Internet dissipated quickly. It was not realistic to think that every web resource could be cataloged like a book or a DVD in a library. Instead the profession and the general public have embraced search engines to retrieve materials sought on the Internet. Although their results are sometimes overwhelming, often full of false drops, and currently lack semantic controls (e.g., synonyms and homographs are not addressed), search engines are excellent for satisfying most general searching needs. Information professionals, researchers, and other conscientious information seekers do not perform in-depth, scholarly research using only search engines, but they may start their preliminary exploration of a topic that way. Search engines, like other more narrowly focused retrieval tools, have an important role to play in information retrieval, and it is anticipated that their scope, power, and effectiveness will grow in the future as more meaningful metadata is added to web content.

Although some believe that organizing the Internet is impossible, the parts that are important for retrieval and for posterity *will* be brought under some organizational control. It is human nature to want to organize, and the principles learned over centuries of organizing print information can be an inspiration, and in some cases can be used alongside newer tools to help organize online information. The current effort to create the Semantic Web, wherein online data are more meaningfully defined and linked to other relevant data for the purpose of more effective discovery of information, is a case in point.

Semantic Web and Linked Data

In the rapidly advancing, technology-centered environment of the third millennium, much of our lives are focused on the Internet. A high volume of the world's information has been transferred to the web, as has much of its metadata. Many of our modern work processes and methods of communication are dependent on its technologies. The Internet is indispensable in our twenty-first-century lives, but as we have already discussed, it is flawed. Machines are not very smart, and they lack intuition. They are not good at inferring connections or understanding context without being fed significant amounts of additional information. Consequently, we are sometimes frustrated with our technology. At this time the World Wide Web and its search engines lack semantic controls. When we search Google for information about spiders on the web (the programs that crawl the Internet looking for new websites), we are also inundated with pages about and pictures of the cobwebby things created by some arachnids to catch their dinner. Synonyms and homographs make searching our current keyword-matching systems an obstacle course of false drops. And, for the most part, users are accustomed to it; they accept it as a part of the search process. Not everyone, however, feels that this is okay.

In 1999 Sir Tim Berners-Lee first described his dream that someday computers would be able to analyze all the data on the web (including transactions between computers and people),[36] and when that became possible, machines talking to machines would bring about a so-called *Semantic Web*. The Semantic Web is a proposed extension to the traditional World Wide Web. Different scholars have expressed its goals in a variety of ways:

- "The Semantic Web will bring structure to the meaningful content of Web pages, creating an environment where software agents roaming from page to page can readily carry out sophisticated tasks for users. . . . [It] will usher in significant new functionality as machines become much better able to process and 'understand' the data that they merely display at present."[37]
- "The general vision of a 'semantic web' can be summarized in a single phrase: *to make the web more accessible to computers*."[38]
- "The goal of the Semantic Web is to move from a Web of Documents to an open inter-connected Web of Data."[39]

What these statements indicate is that the current web is somewhat lacking in structure, semantics (i.e., meaning), and focus. This means that the Semantic Web must shift from identifying and retrieving documents for human consumption *only* to identifying and retrieving the data in those documents for humans *and* for better machine understanding, manipulation, and processing. In many ways, it is an attempt to turn the web into something like a giant global database. So, if we go back to that analogy of the Internet being a library where all the books are dumped on the floor,[40] then the Semantic Web is a library where the books have been ripped apart into individual pages, paragraphs, or even sentences (i.e., chunks of data), so that chunks of data can be clearly identified and relationships between them can be explicitly indicated in order for machines to connect them on a more granular level. The notion of the Semantic Web is not based on the document as a unit; it is more concerned with what is contained within.

There are some important differences between the Semantic Web and the traditional document-based web. At first, in the document-based web, we typically thought of a resource as a web page—a document encoded in HTML that contained information; it was a static, unidirectional transaction. Over time, this model has evolved; web resources are now more than just sites and documents, they include news outlets, commercial services (e.g., Amazon.com), technology products and services, interactive features (e.g., Facebook, Twitter), and so on. On the Semantic Web, however, the term *resource* is defined yet even more broadly; it can refer to websites, online databases, and so on, but *resource* can also refer to real-life objects (e.g., a physical book, a painting, a chalice) as well as people, places, ideas, and so on. In other words, *anything*, electronic or not, can be represented on the Semantic Web as long as there is a *Uniform Resource Identifier* (URI) or an *Internationalized Resource Identifier* (IRI) that can be unambiguously connected to it. "A Uniform Resource Identifier (URI) is a compact sequence of characters that identifies an abstract or physical resource."[41] In other words, it is a character string that helps to identify things. The familiar URL—what we see in a web browser's address bar—is one type of URI. An IRI is similar to a URI but works with an expanded, international character set.

In the traditional document-based web, pages are containers of information. That information is marked up in HTML to indicate some major features (e.g., title, headings, paragraphs) and to denote how the content should be displayed, but very little semantic structure is found. There is little to no indication of what *types* of data are found in a resource, who is responsible for it, what it is about, or that there are thousands of other resources that exist on the same topic (or are connected to it in some other way). The Semantic Web, on the other hand, hopes to supply better semantic structures in the data to improve understandability and to increase connections. This is accomplished by creating additional metadata that describes, in more meaningful ways, the information found in web resources. This means that documents need to be described using metadata in a structured form, so that machines can process that data and get a better "understanding" of the content. For example, using a highly structured format called the *Resource Description Framework* (RDF), a web resource found at www.loc.gov/aba/cataloging/tools/ could be coded to say: the Library of Congress (LC) is the creator, the title is "Resources for Cataloging," and the document is about *cataloging*. In RDF, though, it would use IRIs to communicate that information. These statements would look something like what you see in Figure 1.2. (There are several methods of communicating RDF statements; this is the simplest.) RDF is a structure for communicating metadata on the web; it is described further below (and is explained in more detail in Chapter 5).

Whole documents can be described as seen in Figure 1.2, but so can individual pieces of content *within* documents through the use of additional attributes. To increase better understanding of individual chunks of content, *microdata*[42] (an HTML specification used to nest metadata within the content of web resources) can be embedded into a web page. For example, using semantics defined by Schema.org[43] (a vocabulary designed for marking up electronic content) the following statement could be added to the HTML code of the Library of Congress homepage:

```
01 <https://www.loc.gov/aba/cataloging/tools/> <http://purl.org/dc/terms/creator> <http://id.loc.gov/authorities/names/n78089035> .
02 <https://www.loc.gov/aba/cataloging/tools/> <http://purl.org/dc/terms/title> "Resources for Cataloging" .
03 <https://www.loc.gov/aba/cataloging/tools/> <http://purl.org/dc/terms/subject/> <http://id.loc.gov/authorities/subjects/sh85020816>.
```

Figure 1.2. An Example of RDF.

```
<div itemscope itemtype="http://schema.org/Government
Organization">
    <h1 itemprop="name">Library of Congress</h1>
    <span itemprop="alternateName">@librarycongress</span>
</div>
```

This states that the web resource is for a government agency, its name is *Library of Congress*, it is also known by its twitter user name *@librarycongress*, and the two names are roughly equivalent.

Using RDF structured metadata or embedded microdata, any and all of the following information *could* be added to improve machine understandability of the content:

- LC is the creator of the page;
- LC is a library;
- LC is part of the legislative branch of the United States government;
- *www.loc.gov* is the web address of a document that is created by, but is also about, LC;
- *www.loc.gov* may be used to represent the real-world entity named *Library of Congress*;
- LC has a general contact phone number of (202) 707-5000;
- LC is in Washington, DC (a geographic location, which could *also* be described extensively); and
- LC is specifically located at 101 Independence Ave, SE, Washington, DC 20540.

In today's web of documents, though, the HTML encoding provides little more instruction than the following:

- present the document using HTML;
- identify this particular string of text as the title of the document;
- present these other strings as headings (and, therefore, larger);
- format most content into paragraphs;
- use a particular style sheet for details on how things are to be presented; and
- provide a link to the library catalog where *Search and browse records* is written; and so on.

According to Tom Heath and Christian Bizer, "While most Web sites have some degree of structure, the language in which they are created, HTML, is oriented towards structuring textual documents rather than data. As data is intermingled into the surrounding text, it is hard for software applications to extract snippets of structured data from HTML pages."[44]

For many years, pages on the traditional web provided a limited number of hand-selected links to other web pages. Creators chose to point to other documents that they considered to be relevant to the information they were presenting—generally at the level of the whole resource or a discrete part of it. As the web developed, technologies were developed that allowed links to be automatically generated. A goal of the Semantic Web is to increase technology's abilities to connect an even wider variety of related resources through automatic means. The notion of linked data is the bedrock for the Semantic Web. *Linked data* is an approach to encoding data from a wide range of different sources, and publishing it on the web in such a way that it may be understood by computers as related to the same resource or concept. Linked data

- is machine-readable,
- has meaning explicitly defined,
- is linked to other external data sets, and
- is linked to *from* external data sets.[45]

For example, if two different resources contain metadata statements that say each of those resources is about *cataloging* (with each resource pointing to an IRI for an authorized description of that concept), then the machine could "understand" that those resources are about cataloging, and are, therefore, related to each other. The machine can exploit the connections and infer that if one of these resources were of interest, the other would be of interest to users as well.

With some intellectual effort, the often-tacit relationships between resources represented on the web (e.g., creators, subjects, locations) can be made more obvious, more explicit, and more easily machine actionable through the use of linked data. Linked data uses RDF as its foundation to ensure this clarity in meaning.[46] RDF is a sentence structure for metadata statements, and that structure is referred to as a *triple* because it contains three parts: a subject, a predicate, and an object. For example, a resource could contain an encoded equivalent to this statement:

This web page	is about	cheese.
Subject	**Predicate**	**Object**

Of course, when it has been encoded for machines rather than for human understanding, it is no longer in natural language. Instead it uses IRIs from authoritative sources to replace the subject, the predicate, and whenever possible, the object (although sometimes the object is a *literal*—a fixed string of characters). For example, that same statement would look something more like this:

<http://www .cheese.com>	<http://purl.org/dc /terms/subject>	<http://dbpedia.org /page/Cheese>
Subject	**Predicate**	**Object**

The subject is the web resource being described (www.cheese.com). The predicate explains the relationship between the resource and the object of the statement; it is an IRI for the *Subject* element in the Dublin Core Metadata Element Set.[47] In other words, it says this web resource is *about* the concept represented in the object. The object is an IRI that points to the dbpedia page for the concept *Cheese*. Quite simply, the triple is a machine-understandable way to say: "This web page is about cheese," or "This web page was created by Samantha Jones," or "This resource is titled *I Love You, Cheese*," and so on.

If the Semantic Web's data is marked up more effectively and more thoroughly using triples that identify all the important persons, places, things, ideas, and so forth, computers can use that metadata to make links between related resources. When many resources are linked to the same or equivalent IRIs, those resources are then connected; and all those connected resources are potentially of interest to users searching for materials about a particular topic, by a particular creator, from a particular place or year, and so on. These connections can be traversed automatically, allowing machines to offer a wider variety of, but also more accurately pinpointed, resources as part of search results (i.e., they will not be reliant solely on matching keywords in documents to words in search boxes). This will allow not just better search results but also information collections that can be created by automatic means and updated dynamically without human intervention; this brings together diverse resources that, in many cases, human beings would be unlikely to find. Linked data makes possible the discovery of information about entities that would otherwise have been separated by disassociated means of encoding or by different data silos.

Despite being first described at the beginning of the 21st century,[48] the practical aspects of Semantic Web development are still in their early stages. While its foundations are being laid, there are only a few concrete products to demonstrate proof of concept. The British Broadcasting Corporation (BBC) has gained notice as one of the first organizations to use linked data throughout various parts of its website. For example, at BBC Music,[49] you may search for a page dedicated to your favorite musician. On that page, you will find images, biographical data from Wikipedia, audio and video files from the BBC, links to playlists that contain music by that artist, information about upcoming live performances from Songkick.com (a site dedicated to concert information), recent news stories from the BBC, and links to other sites about the musician (e.g., MusicBrainz sites, official homepages, Facebook pages, Twitter feeds, Instagram profiles, YouTube channels), as well as information about similar artists. The important thing to realize about this example is that machines have done the work; computers have pulled together this collection of data from diverse sources to present users with a wider variety of information.

The current activities in place to embrace linked data may eventually bring a full Semantic Web to fruition, but much work must be completed in order for that to happen. Adding more metadata and microdata to the web is an enormous task. In order for it to succeed, metadata must be created in a more collaborative fashion. In other words, the responsibility for creating the needed metadata must be shared among many, many institutions. The diversity attained by using metadata from multiple sources and multiple perspectives can help to create a more robust picture of the entities described for machines *and* for humans. And the fuller that picture is, the more connections can be made.

Many see the development of linked data and the Semantic Web as the next frontier of information organization, and concrete steps are being taken to help the LIS profession to start on this journey. One such project is the Library of Congress's Bibliographic Framework Initiative (BIBFRAME), which is an attempt to replace the MARC encoding format currently used in catalogs with a linked data approach to representing and exchanging bibliographic data on the web and in the networked world.[50] This project, like many others, is in its infancy and should be watched carefully by anyone interested in information organization. These grassroots efforts, once again, are a clear demonstration of the basic human desire to organize for the common good.

There is great excitement about the potential of the Semantic Web, and the changes that could occur as the result of its maturation. There will be far-reaching effects on information organization and cataloging, but exactly *how* things will change is not completely evident at this time. Although some have expressed hope about this journey, there are still plenty of naysayers who cannot imagine the vision of the Semantic Web actually being manifested.[51]

Digital Collections

The development of the Internet has changed many things. We now have much faster means of communication, instant shopping, news and entertainment on demand, endless pictures of cats, new ways to meet the love of your life, and plenty of other distractions. It has also redefined the notion of space. We now live much of our lives in online spaces—venues to visit for a range of reasons (e.g., to catch up with friends, to attend classes, to watch television, to spy on an ex). One of the more ambitious and beneficent uses of online spaces has been the development of *digital collections*. These online spaces, generally, contain collections of digitized or born-digital resources that have been selected for inclusion for the purposes of access and preservation for posterity. A specific digital collection may be referred to as a *digital library*, a *digital archives*, or an *institutional repository*. These terms are not always precisely defined; what is a *repository* in one institution may be very similar to the *digital archives* of another. What an individual digital collection is called may depend on the nature of the content and the nature of the institution establishing the collection. For example, one would expect to find mostly digitized primary sources (i.e., original documents) in digital archives.

Digital collections are more sophisticated and more accessible than ever before, but they came from humble beginnings. In the 1990s the definition of *digital library* was a matter of debate. There were many different types of experiments and projects labeled as such that would not be considered to be digital libraries today. For example, at the simplest level were collections of links to resources related to a particular subject; sometimes, such collections (really bibliographies) were coordinated among individual librarians at cooperating institutions. To some, the term referred to sites like Project Gutenberg, which in its earliest form was a large collection of "e-texts" to be read online as simply formatted web pages. LIS scholar and educator Candy Schwartz states, "It is an understatement to say that the phrase 'digital library' . . . means different things to different people." As part of a class project, her students found 64 formal and informal definitions of *digital library*.[52] But over time the phrase came to refer to collections in which a site provides digitized information resources with a defined architecture and a service for the retrieval of such resources. Summarizing the definitions that had become predominant by the turn of the 21st century, including those of Schwartz,[53] it can be said that a digital library

- must contain an organized collection of digital resources (it is not a set of hyperlinks to other material);
- is created for a particular audience, group of users, or community;
- takes advantage of technology and human resources (e.g., librarians);
- provides fast and efficient access to digital resources, often without cost (although membership in a community might be required); and
- owns, controls, or has rights to the digital resources distributed by it.

Christine Borgman, an LIS scholar, states: "Digital libraries are an extension, enhancement, and integration both of information retrieval systems and of multiple information institutions, libraries being only one. The scope of digital libraries' capabilities includes not only information retrieval but also creating and using information."[54] She emphasizes that digital libraries are for communities of users and that they are really extensions of the physical places (e.g., libraries, museums, archives, schools) where resources are selected, collected, organized, preserved, and accessed.[55] Today, one example of a digital collection usually found in an academic setting is the *institutional repository*—an online system that collects, manages, disseminates, and preserves digital resources related to the intellectual activity of an academic community.

A successful project that began early in the history of digital libraries and is still ongoing is the American Memory digital collection at the Library of Congress.[56] It began with a pilot program in 1990 and has grown into one of the most well-known digital collections in the world. From the beginning, this digital library has been a collaborative effort, with public, research, and academic libraries, museums, historical societies, and archival institutions digitizing American history collections and making them available on the American Memory site.

Another example of a cooperative project is the Digital Public Library of America (DPLA), planning for which began in 2010. Its stated goal is "bringing together the riches of America's libraries, archives, museums, and cultural heritage sites, and making them freely available to students, teachers, researchers, and the general public."[57] One year after the official launch in April 2013, the DPLA "contained over 7 million digitized cultural heritage items from 1,200 contributing institutions across the United States—up significantly from 2.4 million items and 500 institutions at launch."[58] And, it is, of course, still growing. A similar project, the Europeana Collections launched in 2008, currently has more than 53 million "artworks, artefacts, books, videos and sounds from across Europe."[59] One of the more innovative aspects of the DPLA project is their creation of "a platform that enables new and transformative uses of our digitized cultural heritage. The DPLA's application programming interface (API) and open data can be used by software developers, researchers, and others to create novel environments for learning, tools for discovery, and engaging apps." For example, there are interfaces, linked from the DPLA website, that allow users to search images by color within the collection, to compare search results for two different terms or phrases, to play metadata games (e.g., adding keywords to images in the form of a game), and to find random images in the collection that contain cats.[60]

What might be found in a digital collection? Almost anything can be included as long as it can be presented in digital form. Although some digital collections are very narrowly focused, with limited resource types included, others are more sweeping in scope. For example, in the FDR Presidential Library and Museum's digital collection FRANKLIN, one primarily encounters digitized texts—speeches, correspondence, diaries, press conference transcripts, executive orders, and so on.[61] One can certainly find large amounts of text among most digital collections, but much of the excitement about these spaces is for the inclusion of multimedia resources and complex digital objects. In digital collections like American Memory or in the DPLA, one will find a great variety of resource types. For example, one may find sound files of historic speeches, interviews, lectures, entire long-playing record albums from yesteryear, exotic nature sounds, and so on. Short films, video clips, television programs, political ads, news footage, public service announcements, and the like are not unusual, nor are digitized photographs, drawings, advertisements, and advanced geospatial data renderings. This diversity is now practically *de rigueur* for digital collections.

The first digital collections were custom developments, as there were not yet off-the-shelf software packages to house these early digital projects. For example, the University of California, Berkeley, had several projects that have contributed to digital library innovations. One project created specifications for encoding electronic finding aids for special collections and archives. This project was the basis for what has become Encoded Archival Description (EAD), a key metadata standard in archives.[62] Another project was the Making of America II project, which proposed a standard encoding for digital objects; this evolved into a nationally recognized standard for digital collections, the Metadata Encoding & Transmission Standard (METS).[63]

As the emphasis has shifted from experimentation to mainstream implementation, focus has changed to emphasize standardization, organization, usability, and production of commercially available packages to be used by institutions just entering the digital library arena. Several library automation companies offer digital library solutions. For example, Ex Libris offers DigiTool, a package, working in conjunction with an integrated library system, which includes the means to "facilitate cataloging, managing, sharing, searching, and retrieval of digital collections."[64] Other such products are Portfolio offered by SirsiDynix[65] and Vital from Innovative Interfaces.[66] There is also independent software that institutions can use to build their own digital collections. One such product is Omeka, a project of the Roy Rosenzweig Center for History and New Media at George Mason University.[67] Omeka is free, open-source software that integrates many different document formats but with standardized metadata. "Omeka is a Swahili word meaning to display or lay out wares; to speak out; to spread out; to unpack;"[68] it is a very apt name for a tool that allows libraries, archives, museums, and other cultural heritage institutions to display their digital resources. It was created specifically to be easy to use for those who know nothing about digital libraries but want to create digital collections or exhibits. For an institution with significant technology resources, however, it can be quite a powerful tool, and there is a large community of coders who can provide help as well as add-ons to enhance the system. Other software products for digital collections include ContentDM available from OCLC,[69] DSpace,[70] Fedora,[71] and Digital Commons.[72]

Organization of digital collections is often accomplished using some of the same organizing methods that are used in physical libraries, archives, and museums, but these approaches to organization are supplemented with the use of additional tools and metadata schemas, as well as other controlled vocabularies, taxonomies, encoding formats, and content standards. Approaches to organization can vary from institution to institution and among software packages. As with tangible resources, the important attributes of digital resources are described, access points are chosen, and the resources are analyzed so that subject matter and genre/form can be described. In addition to the metadata generated from those activities, further metadata may be needed to address the resource's digital lifecycle. For example, although users need to know that the resource is text-based and 120 pages in length, they also need to know in which formats the resource is available (e.g., as HTML or PDF). If the resource is downloadable or if it requires software to interact with it, the size and the systems requirements are necessary metadata. If there are restrictions on access or use, these should be stated clearly. The description, in addition to identifying those parties responsible for the creation of the content, might also include those who digitized the resource and when that occurred, as well as who created the metadata, when it was created, and when it was last updated. Some structural metadata might also be needed to help the pieces of complex digital objects fit together to operate smoothly (e.g., so the video can be viewed or the pages of an e-book can be turned). Digital resources can be described at various levels.

To contain the necessary metadata for each resource, some digital collections use established schemas such as METS, DC, Metadata Object Description Schema (MODS), or even full traditional library cataloging (i.e., creating MARC records using RDA, LCSH, and other complementary standards). For example, LC uses both MARC-based and non-MARC metadata in American Memory. Other institutions, however, develop their own approaches, often partly or fully based on one or more previously existing metadata schemas. For example, the DPLA's Metadata Application Profile (MAP) identifies several existing metadata schemas as sources for their set of metadata elements. Rather than "reinventing the wheel," they reused elements from DC, the Europeana Data Model (EDM), the Simple Knowledge Organization System (SKOS), and other schemas. Like many other digital collections, DPLA employs strategies that are compliant with the protocols used by the Open Archives Initiative (OAI) to ensure easy sharing of digital resources and metadata. Projects such as these may also use one or more lists of controlled terminology to describe subjects, names, places, titles, forms, genres, resource types, and so on. For example, American Memory uses controlled names from the LC/NACO Authority File (LCNAF) and subjects from LCSH in its metadata. In short, the organization of digital libraries is being accomplished by using many of the same approaches used in libraries, archives, museums, and other cultural heritage institutions, with tools such as content standards, metadata schemas, Extensible Markup Language (XML) or MARC encoding, controlled vocabularies, user tagging, taxonomies, and so forth. These components are described more fully in later chapters of this book.

Information Architecture

Digital libraries, general websites, online news sources, online banks and bookstores, mobile applications, Facebook, and the like are all examples of online spaces. Just as architects determine the needs of the people who will use a physical space and then design buildings that will serve those needs, so must information architects determine the uses to which information will be put and create paths to finding needed information in online spaces. Information architecture, then, is more than just an aesthetic exercise in website design, although its development and emergence as a discipline is closely tied to web design; it is connected to areas of study within LIS, such as usability and the organization of information, as well as to other disciplines.[73] Andrew Dillon defines *information architecture* (IA) as "the process of designing, implementing and evaluating information spaces that are humanly and socially acceptable to their intended stakeholders."[74] He says he purposely leaves the definition "open so that we cover the organizational, blueprinting, and experience aspects, and allow for IA roles to cover these aspects."[75] Rosenfeld, Morville, and Arango define *information architecture* as:

- The structural design of shared information environments.

- The synthesis of organization, labeling, search, and navigation systems within digital, physical, and cross-channel ecosystems.
- The art and science of shaping information products and experiences to support usability, findability, and understanding.
- An emerging discipline and community of practice focused on bringing principles of design and architecture to the digital landscape.[76]

Then, they jokingly ask, "Were you expecting a single definition? Something short and sweet? A few words that succinctly capture the essence and expanse of the field of information architecture? Keep dreaming!"[77]

There is still disagreement about what is part of IA. Andrew Dillon and Don Turnbull describe some differences in points of view:

> In the absence of formal definition, a line of division has been drawn between two competing views of the field, known generally as the Big IA vs. Little IA perspectives. Big IA is . . . the process of designing and building information resources that are useful, usable, and acceptable. From this perspective IA must cover user experience and even organizational acceptance of the resource. On the other hand, Little IA refers to . . . a far more constrained activity that deals with information organization and maintenance, but does not involve itself in analyzing the user response or the graphical design of the information space. Big IA tends to be seen as top-down, conceiving the full product and its human or organizational impact; Little IA is viewed as more bottom-up, addressing the metadata and controlled vocabulary aspects of information organization, without dealing directly with, and certainly never evaluating formally, the user experience of the resulting space.[78]

Some information architects reject the notion that information architecture is a new approach to the information organization that has been practiced in libraries, archives, and museums for a long time. But the parallels are striking. Librarians have long understood the necessity of organizing information resources in ways that will aid users in gaining access to them as needed. There does appear to be some agreement, however, on the desire to design complete and helpful "information ecologies." To do this, the information architect must

- create online spaces with users' needs, behaviors, and limitations in mind;
- understand the specific context of the site (e.g., the mission, goals, strategies, etc., that are unique to the institution creating the space); and
- organize the online information (e.g., the stuff that users are looking for) logically and clearly to provide easy access to that information.[79]

The process includes designing usable interfaces and navigation systems (e.g., sitemaps, taxonomies, menus) in addition to creating a pleasing overall graphic design. Rosenfeld, Morville, and Arango state, "many people think of website navigation structures when they think of information architecture, and this view isn't entirely off: navigation menus and their ilk are certainly within the remit of what information architecture produces. It's just that you can't get there without having explored the more abstract territory first."[80] In short, the field is dedicated to making information findable and understandable by creating unique and logical information structures in online settings.

Rosenfeld, Morville, and Arango identify the following stages that the process of information architecture must go through: research, strategy, design, implementation, and administration.[81]

- ***Research*** includes a review of background materials; gaining an understanding of the goals and business context; examining the existing information architecture, content, and intended audiences; and finally conducting studies necessary to explore the situation. In this stage, an understanding of the content must be developed. This includes gathering information about: ownership; types, formats, and structures of the content; existing metadata; and the amount or volume of content.
- ***Strategy*** arises from contextual understanding developed in the first phase and defines the top levels of the site's organization and navigation structures, while also considering document types

and the metadata schema. For example, one must develop ideas about how users will access the site's information (e.g., alphabetic, chronological, topical, or task-oriented means).

- *Design* involves creating detailed blueprints, metadata schemas, and the like, to be used by graphic designers, programmers, content authors, and the production team. In this stage, content categories, browsing menus, controlled vocabularies, search functions, and label systems are created.
- *Implementation* is where designs are used in the building, testing, and launching of the site—organizing and tagging documents, troubleshooting, and developing documentation occur in this phase.
- *Administration* involves the continuous evaluation and improvement of the site's information architecture.

The strategy and design stages, in particular, are the ones that require a thorough understanding of the theoretical underpinnings of information organization and the system design that will allow display of results in a logical and usable fashion.

Indexing and Abstracting

Indexing and abstracting are two approaches to distilling information content into an abbreviated, but comprehensive, representation of an information resource. Indexing has a long and shifting tradition in terms of what it is, who has done it, why it is done, and how it is done. The history of abstracting is less volatile, and has evolved in the twentieth and twenty-first centuries into specific formats with targeted audiences.

Indexing

Indexing is the process by which the content of an information resource is analyzed, and the aboutness of that item is determined and expressed in a concise manner. Indexing is also concerned with describing the information resource in such a way that users are aware of the basic attributes of a document, such as author, title, length, and the location of the content. Indexing typically concerns textual items only; although image indexing is a growing area of practice. There are three basic types of indexing: back-of-the-book indexing, database indexing, and web indexing.

In traditional *back-of-the-book indexing,* the index is a list of terms or phrases arranged alphabetically with locator references that make it possible for the user to retrieve the desired content. Language of the indexing terms is typically derived from language of the text—thus the kind of indexing done in this context is referred to as *derived indexing*. A good book index will also include second-level entries (i.e., subheadings), variant entries (i.e., multiple entry points), and cross-references. Book indexing is primarily done by freelance specialists who contract with publishers, although some publishers maintain an in-house indexing staff, or it may be accomplished by the authors themselves.

In *database indexing* (sometimes referred to as *journal* or *periodical indexing*), each database item is represented by a set of descriptor terms and, in some instances, a classification code. Database indexing normally uses a controlled vocabulary or thesaurus from which the indexer selects and assigns the appropriate terms. The scope and number of descriptor terms assigned to an item is determined by the editorial policies of the publisher of the given database, and in-house or specially trained indexers usually perform the indexing.

Web indexing (or *Internet indexing*) is a type of indexing still very much in development, in terms of both its jargon and its actual practice. Currently, web indexing falls into two basic categories: (1) back-of-the-book style indexing—often referred to as *A–Z Indexing*—which uses encoded index links within the website, and (2) *search engine indexing*—more accurately described as the automatic indexing of websites. In search engine indexing, websites are searched based on user query terms, an index of the words found and where they were found is maintained, and future searching on these same queries uses the saved indexes. Because of the varieties of web indexing, *who* does web indexing is a question that applies primarily to the A–Z style of web indexing. Freelance contractors, often book indexers who have expanded their repertoire of services, create A–Z web indexes. Different types of indexes are described more fully in Chapter 3.

A number of software tools have been created to generate indexes. They use a variety of techniques, the efficacy of which depends upon variables such as cost, time constraints, type and size of files to be indexed, and individual preferences. The American Society for Indexing (ASI) lists several types of tools used for indexing.[82] These include:

- Standalone or dedicated tools, which allow the creation of back-of-the-book indexes from page-numbered galleys;
- Embedded indexing tools, which allow insertion of index entries as invisible text in electronic files;
- Tagging and keywording tools, which allow indexing codes (instead of invisible text) to be embedded in electronic text and allow creation of hard-coded jumps, similar to web links; these rely on the words used by the author of the text, not indexers' concepts;
- Automated indexing software, which accompanies most word-processing software and builds concordances or word lists directly from texts using the language of the authors (again, not true indexes that include key concepts that may not use the author's words);
- Free-text and weighted-text searching tools, which allow the assignment of values to words and phrases.

Abstracting

Abstracting is a process that consists of analyzing the content of an information resource and then writing a succinct summary or synopsis of that work. Typically, abstracting is done for an academic publication or a professional journal. The length, style, and amount of detail in an abstract may vary depending on its intended audience. Generally, an abstract is not a review of the work, nor does it evaluate or interpret the work that is being abstracted, although *critical* abstracts do include some evaluative text. Although it contains key words and concepts found in the larger document, the abstract is an original text rather than an excerpted passage. There are several types of abstracts:

- **Indicative**: descriptive of the content, but without providing results or outcomes (also can be referred to as a *descriptive abstract*);
- **Informative**: summative with the results or outcomes emphasized;
- **Critical**: condensed critical review (this type is fairly uncommon);
- **Structured**: non-narrative in format; it includes specific factors, such as objectives, methods, results, etc.;
- **Modular**: includes five discrete sections—citation, annotation, indicative abstract, informative abstract, and critical abstract.

Technically, an abstract is the summary text; in practice, a formal abstract consists of the title and citation of the abstracted work and the summary text.

Abstracting is done by both authors and specially trained information professionals. Scholarly journals often require that an abstract accompany the articles that authors submit for publication. The rubrics provided by journal publishers can be inconsistent or vague, and the quality of published abstracts can suffer as a consequence. Abstracts are also written by professionals, who are either in-house or are contracted by the publishers. Editorial policies are employed to guide abstractors in this instance; policies are not all the same but are designed in response to specific audience needs.

Abstracts have a number of uses in information organization and retrieval. Users needing to stay abreast of a field or given topic can do so by reviewing abstracts published in that area. Beginning researchers or researchers seeking to master a given area of literature find that reviewing abstracts instead of full texts saves time. Abstracts aid in the decision of which articles need to be read in full versus which can be skimmed or skipped altogether. In a related fashion, because it is sometimes the practice to publish English-language abstracts for non-English-language articles, the user can decide from reading the abstract if it is cost- and time-effective to have a translation made of a full article. Librarians and other information professionals find that the use of abstracts assists in the speed and utility of patron literature searches. Database indexers, who typically index only using the title and abstract of a text, require that the abstract be well written and accurate.

Records Management

Records management is the terminology applied to the control and disposition of records created in offices and other administrative settings. It has its roots in the office filing systems that developed throughout the twentieth century. These systems have been highly affected by developments in technology—typewriters, photocopiers, and computers (starting with sorters and collators). The use of computers in this context has sometimes been referred to as *data administration*. Records management systems have a strong relationship with archives, as that is where an organization's records may be deposited when their active operating life has passed and they have become inactive records.

As was true in other parts of our society, records management originally involved the keeping, filing, and maintaining of paper records. It was a simpler time but also a frustrating time, because usually only one copy of a record was filed in only one place. The file labels of one records manager were not necessarily logical to the next. As information began being entered and stored in electronic files, access points (the file labels) became invisible. This was not an immediate problem as long as the people who developed the electronic files documented what was contained in them. The situation became more complicated when powerful personal computers began to allow persons to store and file their own information on their desktops. A problem of continuity developed when these personal files were abandoned.

For many years various operations were automated, each with its own system. For example, payroll, general ledger, accounts payable, inventories, and other such systems were automated separately. During the 1990s integration of these systems took place with the result that the systems had many redundant data fields with little documentation of their content. These fields seemed to be meant to contain the same information, but what was actually there was often different (e.g., name given in full in the payroll file, but middle name shortened to an initial in the faculty file).

In 2001 the International Organization for Standardization (ISO) published a standard for records management.[83] It defines records management as the "field of management responsible for the efficient and systematic control of the creation, receipt, maintenance, use and disposition of records."[84] Further, in describing records systems characteristics, the standard states:

A records system should

a) routinely capture all records within the scope of the business activities it covers,
b) organize the records in a way that reflects the business processes of the records' creator,
c) protect the records from unauthorized alteration or disposition,
d) routinely function as the primary source of information about actions that are documented in the records, and
e) provide ready access to all relevant records and related metadata.[85]

There have now been developed a number of commercially available records management systems that track and store records, provide security and auditing functions, have content management and user identity modules, and more. Records management systems are a growing industry as corporate settings (as well as college and university, governmental, and other more traditional institutional settings) engage in records management activities and seek technological solutions to long-standing data management problems.

Records managers have dealt with the information explosion by using principles of information organization. The units that need to be organized in the administrative environment are such things as directories, files, programs, and, at another level, such things as field values. Organization can be by system (e.g., payroll, budget) or by type of record (e.g., personal names, registration records). Records managers must keep track of information that crosses system boundaries (e.g., personal names cross boundaries when the same names are entered into several different files). There must be methods for handling concepts that have the same names but different purposes (e.g., the concept of *part time* can have different definitions in a university depending upon whether one is talking about payroll, faculty, graduate students, or undergraduate students).

Personal Information Management*

Personal information management is defined as "the practice and the study of the activities a person performs in order to acquire or create, store, organize, maintain, retrieve, use, and distribute the information needed to complete tasks and fulfill various roles and responsibilities."[86] Because personal information management is directly connected to an individual's everyday life, there has been growing interest in the development of tools and devices that facilitate personal information management, as well as more investigation of how people manage their personal information.

People organize their personal information in various formats—including paper documents, emails, photos, music albums, recipes, and different types of digital files—in their personal spaces such as offices and homes. When organizing personal information, people have different habits and processes. For instance, some people have neatly organized offices while others' offices are messy with piles of paper documents.[87] In the case of organizing digital files, there have been ongoing debates on the necessity of organizing them into folders, particularly since people can now simply search for them in personal devices. However, many people report that they still organize digital files into folders; the act of organizing files has more functions than just finding specific items. These functions include reminding people of tasks and helping them further understand the relationships among information items. In addition, a search function can be less useful when there are a number of files with similar names or the exact key words cannot be recalled.

There are many factors that influence organization decisions, such as where and how to organize items. Primary factors include use/purpose of the information item, format of the information item, and topic of the information item.[88] Organizing personal information can be challenging because it involves various decisions that need to be made based on the future use of, need for, interest in, and value of the information, all of which can be hard to predict (and can change easily). Today, two issues also make personal information organization even more challenging. These are *information overload*, which is when a person receives more information than can be processed, and *information fragmentation*, which is having information items scattered across multiple personal devices and tools in different formats.[89] However, regardless of the format, effectively organizing personal information items definitely facilitates finding and using information efficiently, which can increase an individual's productivity.

Knowledge Management

Everyone has heard the phrase "Knowledge is power." Originally, the phrase applied to individuals and implied that persons who increased their knowledge would be able to increase their power in society. During the 1980s it came to be understood that the same thing applied to organizations. At that time, there was much downsizing of organizations in order to reduce overhead and increase profits. In the process, it became obvious that the organizations lost important knowledge as employees left and took their accumulated years of knowledge with them. In the same period there was much technological development that was seen at first as a way to save costs by replacing human workers. Again, though, the knowledge held and applied by humans was not all replaced by machines. For an organization to survive, knowledge is brought to bear in the challenges the organization faces. Management of that knowledge increases its power.

The idea of passing on knowledge gained in a work setting has existed for centuries. Apprentices learned various trades by working alongside masters. Children often followed parents into family businesses. More recently, there have been people known as *mentors.* Also, a person leaving a job is often asked to train the replacement person before leaving.

Knowledge management (KM) is the process of capturing, developing, sharing, and using organizational information to make good, well-informed decisions. This concept came into being as an attempt to capture employees' knowledge with advanced technology so that the knowledge could be stored and shared easily. As people became overwhelmed with the increased availability of information through rapid technological developments, knowledge management took on the additional role of coping with the explosion of information. In the KM context, the process comprises three major components: people, processes, and technology.

Managing knowledge requires a definition of *knowledge,* a concept that has been discussed by philosophers for years without complete resolution. It has been characterized in several ways—for example, as residing in people's minds rather than in any stored form; as being a combination of information, context, and experience; as being that which represents shared experience among groups and communities; or, as a high value form of information that is applied to decisions and actions. R. D. Stacy makes the following observation:

> Knowledge is not a "thing," or a system, but an ephemeral, active process of relating. If one takes this view then no one, let alone a corporation, can own knowledge. Knowledge itself cannot be stored, nor can intellectual capital be measured, and certainly neither of them can be managed.[90]

However, Rosenfeld, Morville, and Arango posit, "Knowledge managers develop tools, processes, and incentives to encourage people to share" what they know.[91]

Dave Snowden notes that knowledge management started in 1995 with the popularization of ideas about *tacit knowledge* versus *explicit knowledge* put forward by Ikujiro Nonaka and Hirotaka Takeuchi.[92] Nonaka and

Takeuchi postulated that tacit knowledge is hidden, residing in the human mind, and cannot be easily represented via electronics; but it can be made explicit to the degree necessary to accomplish a specific innovation.[93] They described a spiral process of sharing tacit knowledge with others through socializing, followed by listeners internalizing the knowledge, and then new knowledge being created, in turn, to be shared. Snowden says that it does not follow that all knowledge in people's minds could or should be made explicit. Often, the knowledge that can be made explicit is just the tip of the iceberg. However, early knowledge management programs "attempted to disembody all knowledge from its possessors to make it an organizational asset."[94] Software programs were created and are being used for this purpose. For example, Knowledge Base Software from Novo Solutions claims that it can provide a training tool for new employees, centralize and retain employee knowledge, and create and update categorized and searchable knowledge management articles, among other things.[95] Collaboration Solutions software (formerly known as Lotus) from IBM claims to create, organize, share, and manage business content in order to provide the right information effectively and efficiently to those who need it, and to "gather and exchange information through professional networks and build communities of experts to help execute tasks faster."[96]

Naresh Agarwal and Md. Anwarul Islam provide a list of technology and non-technology tools and mechanisms for implementing different phases of the knowledge management cycle (e.g., knowledge capture and creation, knowledge sharing and transfer, and knowledge application and use). They put forth a number of considerations about the place of technology in knowledge management:

- A single set of tools cannot be mandated because every organization and its employees will need to decide for themselves which tools and technologies they find easy to use and useful to their current needs.
- Technology tools keep changing, so there cannot be a permanent set of recommendations that will hold true over time. What will remain consistent is the need for knowledge creation, sharing, and use in organizations.
- An organization needs to factor in the cost of adopting any particular set of tools or technology (i.e., buying/licensing and the cost of maintaining).
- Technology is not the most important component in KM implementations. Knowledge management is about people, not about tools and technology. Technology is needed to support people's needs, and not the other way around.[97]

M. C. Vasudevan, Murali Mohan, and Amit Kapoor observe that knowledge management in an enterprise often involves

- identifying, selecting, and cataloging information resources that are pertinent to the enterprise's needs;
- identifying information flow patterns among individuals and among groups (e.g., finding out who asks what questions, learning what information is obtained from which sources, determining what types of information are not easily available or accessible); and
- designing and developing user-friendly systems for accessing the enterprise's knowledge base.[98]

Core issues of concern to people in the information organization business are those of describing, classifying, and retrieving what has been stored. In the context of knowledge management, this means that the organization's knowledge must be sorted out, labeled (i.e., described), and categorized into different subjects or groups (i.e., a taxonomy) if it is to be retrieved when needed. In their study of knowledge management in consulting firms, Ling-Ling Lai and Arlene G. Taylor found that the firms required creation of a knowledge piece at the end of each project. Further, each organization had a template with attributes and facets appropriate to describing both tacit and explicit knowledge gained during the course of a consultation. The researchers observed that the actions of capturing tacit knowledge and then describing it are much like the process of organizing information in libraries: "Essentially, descriptive cataloging and subject cataloging (in LIS terminology) are achieved when consultants work on describing a knowledge piece by completing a standardized template with a number of attributes, and further when they use facets to categorize the knowledge piece. Whether it is called facet analysis, tagging, or providing metadata, the core meaning of cataloging and classification exists in consulting firms as well as in libraries."[99]

CONCLUSION

This chapter discusses basic needs to organize, defines information organization, and presents an overview of a number of different kinds of organizing contexts and environments. While there are differences among these environments, there are also many points of convergence. All of the contexts and environments are interested in describing resources for retrieval and posterity purposes, providing access to resources and helping users to select what is most appropriate for their needs, helping users to understand and explore the information they encounter, analyzing content and describing it consistently, using categories in beneficial ways, and so on.

The following chapters discuss in more detail the processes that have been developed for information organization, those that are being worked on, and the issues that affect their implementation. But first, in the next chapter, a historical look at the development of organizing processes through a number of centuries serves to give us a perspective on where we have been, where we are now, and how far we might go.

Some Important Terms in This Chapter

Aboutness
Abstract
Abstracting
Access Point
Accession Record
Acquisitions
Approval Plan
Authority Control
Back-of-the-book Indexing
Bibliographic Control
Bibliographic Record
Call Number
Cataloging
Collocate
Collocation
Conceptual Analysis
Container List
Content Standard
Cooperative Cataloging
Copy Cataloging
Creator
Cutter Number
Data
Descriptive Cataloging
Digital Collection
Explicit Knowledge
Explore
Facet
Find
Finding Aid
Gifts and Exchanges
Granularity
Identify
Index
Indexing
Information
Information Architecture
Information Fragmentation
Information Organization
Information Overload
Information Resource
Integrated Library System
Interlibrary Loan
Inventory
Knowledge
Knowledge Management
Knowledge Organization System
Library Services Platform
Linked Data
Metadata
Microdata
Obtain
Organize
Original Cataloging
Original Order
Patron-driven Acquisitions
Personal Information Management
Provenance
Public Services
Records Management
Register
Resource
Respect des Fonds
Retrieval Tool
Search Engine Optimization
Select
Selection (Collection Development)
Semantic Web
Series (Archives)
Subject Cataloging
Tacit Knowledge
Technical Services
Title
Translation
Triple
Union Catalog
Work

Some Important Acronyms in This Chapter

AACR:	*Anglo-American Cataloging Rules*
AMC:	Archival and Manuscript Control
API:	Application Program Interface
ASI:	American Society for Indexing
CDWA:	*Categories for the Description of Works of Art*
CCO:	*Cataloging Cultural Objects*
DACS:	*Describing Archives: A Content Standard*
DC:	Dublin Core
DDC:	*Dewey Decimal Classification*
DPLA:	Digital Public Library of America
EAD:	Encoded Archival Description
EDM:	Europeana Data Model
FRAD:	*Functional Requirements for Authority Data*
FRBR:	*Functional Requirements for Bibliographic Records*
HTML:	Hypertext Markup Language
IA:	Information Architecture
IFLA:	International Federation of Library Associations and Institutions
ILS:	Integrated Library System
IRI:	Internationalized Resource Identifier
ISO:	International Organization for Standardization
KOS:	Knowledge Organization System
LC:	Library of Congress
LCC:	*Library of Congress Classification*
LCNAF:	LC/NACO Authority File (previously Library of Congress Name Authority File)
LCSH:	*Library of Congress Subject Headings*
LIS:	Library and Information Science
LSP:	Library Services Platform
MAP:	Metadata Application Profile
MARC:	MAchine-Readable Cataloging
MeSH:	*Medical Subject Headings*
MODS:	Metadata Object Description Schema
NACO:	Name Authority Cooperative Program
OAI:	Open Archives Initiative
OCLC:	Online Computer Library Center
RDA:	*Resource Description & Access*
RDF:	Resource Description Framework
SKOS:	Simple Knowledge Organization System
TGM:	*Thesaurus for Graphic Materials*
URI:	Uniform Resource Identifier
URL:	Uniform Resource Locator
XML:	Extensible Markup Language

NOTES

All URLs accessed June 2017.

1. "Organize," Merriam-Webster.com, http://www.merriam-webster.com/dictionary/organize.

2. Steven Harnad, "To Cognize is to Categorize: Cognition is Categorization," in *Handbook of Categorization in Cognitive Science,* eds. Henri Cohen and Claire Lefebvre (Amsterdam: Elsevier, 2005), 20–45.

3. Lisa Frederick, "Are You a Filer or a Piler?" HGTV, http://www.hgtv.com/remodel/interior-remodel/are-you-a-piler-or-a-filer; Mark Lansdale, "The Psychology of Personal Information Management," *Applied*

Ergonomics 19, no. 1 (1988): 55–66; Thomas W. Malone, "How Do People Organize Their Desks? Implications for the Design of Office Information Systems," *ACM Trans Office Info Systems* 1, no. 1 (1983): 99–112.

4. International Federation of Library Associations and Institutions, IFLA Study Group, *Functional Requirements for Bibliographic Records* [FRBR]: *Final Report* (Munich: Saur, 1998), 8, http://www.ifla.org/publications/functional-requirements-for-bibliographic-records.

5. Pat Riva, Patrick Le Boeuf, and Maja Žumer, "IFLA Library Reference Model: A Conceptual Model for Bibliographic Information," August 2017 (The Hague: IFLA, 2017), 15–16,

6. Clifford Stoll, *Silicon Snake Oil: Second Thoughts on the Information Highway* (New York: Doubleday, 1995), 193.

7. Ronald Hagler, *The Bibliographic Record and Information Technology,* 3rd ed. (Chicago: American Library Association, 1997), 13.

8. For a more robust exploration of all aspects of technical services in libraries, please see G. Edward Evans, Sheila S. Intner, and Jean Weihs, *Introduction to Technical Services*, 8th ed. (Santa Barbara, CA: Libraries Unlimited, 2011).

9. To learn more in-depth information about collection development and acquisitions, please consult Peggy Johnson, *Fundamentals of Collection Development and Management*, 3rd ed. (Chicago: American Library Association, 2014) and Frances C. Wilkinson, Linda K. Lewis, and Rebecca L. Lubas, *The Complete Guide to Acquisitions Management*, 2nd ed. (Santa Barbara, CA: Libraries Unlimited, 2015).

10. Daniel N. Joudrey, Arlene G. Taylor, and David P. Miller, *Introduction to Cataloging and Classification*, 11th ed. (Santa Barbara, CA: Libraries Unlimited, 2015), 15.

11. *RDA: Resource Description & Access*, developed in a collaborative process led by the Joint Steering Committee for Development of RDA (Chicago: American Library Association, 2010). Also available as *RDA Toolkit: Resource Description & Access* (Chicago: American Library Association, 2010), http://www.rdatoolkit.org/.

12. *Anglo-American Cataloguing Rules, Second Edition, 2002 Revision,* prepared under the direction of the Joint Steering Committee for Revision of AACR (Ottawa: Canadian Library Association; Chicago: American Library Association, 2002 [and earlier editions]).

13. IFLA, FRBR, http://www.ifla.org/publications/functional-requirements-for-bibliographic-records.

14. International Federation of Library Associations and Institutions (IFLA) Working Group on Functional Requirements and Numbering of Authority Records (FRANAR), *Functional Requirements for Authority Data: A Conceptual Model*, approved by the Standing Committees of the IFLA Cataloguing Section and IFLA Classification and Indexing Section, 2009, amended and corrected through July 2013, http://www.ifla.org/publications/functional-requirements-for-authority-data.

15. *Library of Congress Subject Headings,* 13th–35th eds. (Washington, DC: Library of Congress, Office for Subject Cataloging Policy, 1990–2013). See also "Library of Congress Subject Heading PDF Files," Library of Congress, http://www.loc.gov/aba/cataloging/subject/.

16. National Library of Medicine, "Medical Subject Headings (MeSH)," http://www.nlm.nih.gov/mesh/.

17. For information about LCC, see Library of Congress Cataloging and Acquisitions, "Classification and Shelflisting," http://www.loc.gov/aba/cataloging/classification.

18. Melvil Dewey, *Dewey Decimal Classification and Relative Index*, 23rd ed., eds. Joan S. Mitchell et al. (Dublin, OH: OCLC, 2011). DDC, Dewey, Dewey Decimal Classification, and WebDewey are registered trademarks of OCLC.

19. Library of Congress, Network Development and MARC Standards Office, "MARC Standards," http://www.loc.gov/marc/.

20. Charles A. Cutter, *Rules for a Dictionary Catalog,* 4th ed. (Washington, DC: Government Printing Office, 1904; reprint, London: The Library Association, 1962), 12.

21. Ibid., 10.

22. “EAD: Encoded Archival Description,” Library of Congress, http://www.loc.gov/ead/.

23. Patricia Harpring, “Categories for the Description of Works of Art,” eds. Murtha Baca and Patricia Harpring; revised March 2014 by Patricia Harpring (Los Angeles: J. Paul Getty Trust, 2014), http://www.getty.edu/research/publications/electronic_publications/cdwa/.

24. Library of Congress, Prints & Photographs Division, “Thesaurus for Graphic Materials,” http://www.loc.gov/pictures/collection/tgm/index/terms/.

25. The Getty Research Institute, “Getty Vocabularies,” http://www.getty.edu/research/tools/vocabularies. This site includes links to: “The Art & Architecture Thesaurus” (AAT), “The Getty Thesaurus of Geographic Names” (TGN), “The Cultural Objects Name Authority” (CONA), and “The Union List of Artist Names” (ULAN).

26. *Iconclass: An Iconographic Classification System* (Amsterdam: North-Holland Pub. Co., 1974–1981), http://www.iconclass.nl.

27. Murtha Baca, Patricia Harpring, Elisa Lanzi, Linda McRae, and Ann Whiteside, on behalf of the Visual Resources Association, *Cataloging Cultural Objects: A Guide to Describing Cultural Works and Their Images* (Chicago: American Library Association Editions, 2006), http://cco.vrafoundation.org/.

28. Artstor, http://www.artstor.org.

29. OCLC, “WorldCat,” https://www.worldcat.org/.

30. Joshua Quittner quoting Ed Krol, “Getting Up to Speed on the Computer Highway,” *Newsday*, November 3, 1992, 51; cited in Arlene G. Taylor, “The Information Universe: Will We Have Chaos or Control?” *American Libraries* 25, no. 7 (1994): 629.

31. Norman Oder, “Cataloging the Net: Can We Do It?” *Library Journal* 123, no. 16 (1998): 47–51.

32. “INFOMINE: Scholarly Internet Resource Collections,” University of California, Riverside, https://web.archive.org/web/20160311214940/http://library.ucr.edu/view/infomine; Danny Sullivan, “The Yahoo Directory—Once The Internet’s Most Important Search Engine—Is To Close,” *Search Engine Land*, http://searchengineland.com/yahoo-directory-close-204370.

33. “SciCentral,” http://www.scicentral.com/index.html.

34. Louis Rosenfeld, Peter Morville, and Jorge Arango, *Information Architecture: For the Web and Beyond,* 4th ed. (Sebastopol, CA: O’Reilly, 2015), 12.

35. Dublin Core Metadata Initiative, “DCMI Home,” http://dublincore.org/.

36. Tim Berners-Lee with Mark Fischetti, *Weaving the Web: The Original Design and Ultimate Destiny of the World Wide Web by Its Inventor* (New York: Harper-Collins, 1999), chap. 12.

37. Tim Berners-Lee, James Hendler, and Ora Lassila, “The Semantic Web,” *Scientific American* 284, no. 5 (2001): 34.

38. Grigoris Antoniou, Paul Groth, Frank van Harmelen, and Rinke Hoekstra, *A Semantic Web Primer*, 3rd ed. (Cambridge, MA: MIT Press, 2012), 1. [Italics added.]

39. Florian Bauer and Martin Kaltenböck, *Linked Open Data: The Essentials* (Vienna, Austria: DGS, 2012), 25, http://www.reeep.org/LOD-the-Essentials.pdf.

40. Quittner, “Getting Up to Speed,” 51.

41. Tim Berners-Lee, Roy Fielding, and Larry Masinter, “Uniform Resource Identifier (URI): Generic Syntax,” January 2005, http://tools.ietf.org/html/rfc3986.

42. Web Hypertext Application Technology Working Group (WHATWG), “Microdata,” *HTML Living Standard*, https://html.spec.whatwg.org/multipage/microdata.html.

43. Schema.org is a project to help “improve the web by creating a structured data markup schema supported by major search engines.” From “About Schema.org,” https://schema.org/docs/faq.html.

44. Tom Heath and Christian Bizer, *Linked Data: Evolving the Web into a Global Data Space* (San Rafael, CA: Morgan & Claypool, 2011), 2.

45. Christian Bizer, Tom Heath, and Tim Berners-Lee, “Linked Data: The Story So Far,” *International Journal on Semantic Web and Information Systems* 5, no. 3 (2009): 2.

46. See Chapter 5 for more on RDF.

47. Dublin Core Metadata Initiative, “Dublin Core Metadata Element Set, Version 1.1,” http://dublincore.org/documents/dces/.

48. Berners-Lee, Hendler, and Lassila, “The Semantic Web,” 34–43.

49. BBC, “Music,” http://www.bbc.co.uk/music.

50. Library of Congress, “Bibliographic Framework Initiative,” https://www.loc.gov/bibframe/.

51. Seth Grimes, “Semantic Web Business: Going Nowhere Slowly,” *InformationWeek*, January 7, 2014, http://www.informationweek.com/software/information-management/semantic-web-business-going-nowhere-slowly/d/d-id/1113323.

52. Candy Schwartz, “Digital Libraries: An Overview,” *Journal of Academic Librarianship* 26, no. 6 (2000): 385.

53. Ibid.

54. Christine L. Borgman, *From Gutenberg to the Global Information Infrastructure: Access to Information in the Networked World* (Cambridge, MA: MIT Press, 2000), 48.

55. Ibid., 42.

56. Library of Congress, “American Memory,” https://memory.loc.gov/ammem/index.html.

57. Digital Public Library of America (DPLA), “Strategic Plan,” 2, https://dp.la/info/wp-content/uploads/2015/01/DPLA-StrategicPlan_2015-2017-Jan7.pdf.

58. Ibid., 3.

59. Europeana Collections, “Home,” http://www.europeana.eu/portal/.

60. DPLA, “Apps Library,” https://dp.la/apps.

61. Franklin D. Roosevelt Presidential Library and Museum, “FRANKLIN,” http://www.fdrlibrary.marist.edu/archives/collections/franklin/.

62. “EAD: Encoded Archival Description,” Library of Congress, http://www.loc.gov/ead/.

63. “METS: Metadata Encoding & Transmission Standard,” Library of Congress, http://www.loc.gov/standards/mets/.

64. Ex Libris, “DigiTool,” http://www.exlibrisgroup.com/category/DigiToolOverview.

65. SirsiDynix, “Portfolio,” http://www.sirsidynix.com/products/portfolio.

66. Innovative Interfaces, “Vital,” https://www.iii.com/products/digital-asset-management/.

67. Omeka, http://omeka.org/.

68. “Project,” Omeka, http://omeka.org/about/.

69. “CONTENTdm,” OCLC, http://www.oclc.org/en-US/contentdm.html.

70. DSpace, http://www.dspace.org/.

71. Fedora, http://fedora-commons.org/.

72. Digital Commons, http://www.bepress.com/products/digital-commons/

73. Andrew Dillon and Don Turnbull, "Information Architecture," in *Encyclopedia of Library and Information Sciences*, 3rd ed., eds. Marcia J. Bates and Mary Niles Maack (New York: Taylor & Francis, 2009), 2362.

74. Andrew Dillon, "Information Architecture in JASIST: Just Where Did We Come From?" *Journal of the American Society for Information Science and Technology* 53, no. 10 (2002): 821, https://www.ischool.utexas.edu/~adillon/Journals/IA%20ASIST%20intro.pdf.

75. Ibid.

76. Rosenfeld, Morville, and Arango, *Information Architecture*, 24.

77. Ibid.

78. Dillon and Turnbull, "Information Architecture," 2362.

79. Rosenfeld, Morville, and Arango, *Information Architecture*, 31–34.

80. Ibid., 20.

81. Ibid., 311–15.

82. American Society for Indexing, "Software," http://www.asindexing.org/reference-shelf/software/.

83. International Organization for Standardization, *Information and Documentation: Records Management*, ISO 15489, Pt. 1. General; Pt. 2. Guidelines (Geneva: ISO, 2001).

84. Ibid., Pt. 1, 3.

85. Ibid., Pt. 1, 8.

86. William Jones, "Personal Information Management," in *Annual Review of Information Science and Technology*, Volume 41 (Medford, NJ: Information Today, 2007), 453.

87. Malone, "How Do People Organize Their Desks?," 104.

88. Deborah K. Barreau, "Context as a Factor in Personal Information Management Systems," *Journal of the American Society for Information Science* 46, no. 5 (1995), 333; Deborah K. Barreau, "The Persistence of Behavior and Form in the Organization of Personal Information," *Journal of the American Society for Information Science and Technology* 59, no. 2 (2008): 315–16; Barbara Kwasnik, "The Influence of Context on Classificatory Behavior," (PhD diss., Rutgers, 1989), 85–86; and Kyong Eun Oh, "The Process of Organizing Personal Information," (PhD diss., Rutgers, 2013), 222–24.

89. William Jones, *Keeping Found Things Found: The Study and Practice of Personal Information Management* (San Francisco: Morgan Kaufmann, 2008), 391–92.

90. R. D. Stacy, *Complex Responsive Processes in Organizations: Learning and Knowledge Creation* (New York: Routledge, 2001), as quoted in Dave Snowden, "Complex Acts of Knowing: Paradox and Descriptive Self-Awareness," *Bulletin of the American Society for Information Science and Technology* 29, no. 4 (April–May 2003): 24.

91. Rosenfeld, Morville, and Arango, *Information Architecture,* 25.

92. Snowden, "Complex Acts of Knowing," 23.

93. Ikujiro Nonaka and Hirotaka Takeuchi, *The Knowledge-Creating Company* (Oxford: Oxford University Press, 1995).

94. Snowden, "Complex Acts of Knowing," 23.

95. Novo Solutions, "Knowledge Base Software," http://novosolutions.com/knowledge-base-software/.

96. IBM, "IBM Collaboration Solutions," http://www-01.ibm.com/software/lotus/.

97. Naresh K. Agarwal and Md. Anwarul Islam, "Knowledge Management Implementation in a Library: Mapping Tools and Technologies to Phases of the KM Cycle," *VINE: The Journal of Information and*

Knowledge Management Systems 44, no. 3 (2014): 322–44.

98. M. C. Vasudevan, Murali Mohan, and Amit Kapoor, "Information System for Knowledge Management in the Specialized Division of a Hospital," in *Knowledge Organization, Information Systems and Other Essays: Professor A. Neelameghan Festschrift,* eds. K. S. Raghavan and K. N. Prasad (Chennai: Ranganathan Centre for Information Studies, 2006), 247–61. [On page 248, the authors acknowledge indebtedness for their ideas to: Hsinchun Chen, *Knowledge Management Systems: A Text Mining Perspective* (Tucson, AZ: Knowledge Computing Corp., 2001); and L. J. Haravu, *Knowledge Management: Paradigms, Challenges and Opportunities* (Bangalore: Sarada Ranganathan Endowment for Library Science, 2002).]

99. Ling-Ling Lai and Arlene G. Taylor, "Knowledge Organization in Knowledge Management Systems of Global Consulting Firms," *Cataloging & Classification Quarterly* 49, no. 5 (2011): 387–407.

SUGGESTED READINGS

General Information Organization and Metadata

Glushko. Robert J., ed. *The Discipline of Organizing*. Cambridge, MA: MIT Press, 2013.

Hider, Philip. *Information Resource Description: Creating and Managing Metadata.* Chicago: ALA Editions, 2013.

Pomerantz, Jeffrey. *Metadata*. Cambridge, MA: MIT Press, 2015.

Weinberger, David. *Everything Is Miscellaneous: The Power of the New Digital Disorder*. New York: Times Books, 2007.

Zeng, Marcia Lei, and Jian Quin. *Metadata*. 2nd ed. Chicago: ALA Neal-Schuman, 2016.

Organization of Information in Libraries

Chan, Lois Mai, and Athena Salaba. *Cataloging and Classification: An Introduction*. 4th ed. Lanham, MD: Rowman & Littlefield, 2016.

Intner, Sheila S., and Jean Weihs. *Standard Cataloging for School and Public Libraries*. 5th ed. Santa Barbara, CA: Libraries Unlimited, 2015.

Joudrey, Daniel N., Arlene G. Taylor, and David P. Miller. *Introduction to Cataloging and Classification*. 11th ed. Santa Barbara, CA: Libraries Unlimited, 2015.

Taylor, Arlene G., and Daniel N. Joudrey. "Cataloging." In *Encyclopedia of Library and Information Sciences,* 3rd ed., edited by Marcia J. Bates and Mary Niles Maack. New York: Taylor & Francis, 2009.

Organization of Information in Archives

Bastian, Jeannette, Megan Sniffin-Marinoff, and Donna Webber. *Archives in Libraries: What Librarians and Archivists Need to Know to Work Together*. Chicago: Society of American Archivists, 2015.

Craven, Louise, ed. *What Are Archives? Cultural and Theoretical Perspectives: A Reader*. Burlington, VT: Ashgate, 2008.

Millar, Laura A. *Archives: Principles and Practices*. New York: Neal-Schuman, 2010.

Roe, Kathleen D. *Arranging & Describing Archives & Manuscripts.* Chicago: Society of American Archivists, 2005.

Organization of Information in Museums

Bierbaum, Esther Green. "Records and Access: Museum Registration and Library Cataloging." *Cataloging & Classification Quarterly* 9, no. 1 (1988): 97–111.

Buck, Rebecca A., and Jean Allman Gilmore, eds. *MRM5: Museum Registration Methods.* 5th ed. Washington, DC: American Association of Museums, 2010.

Latham, Kiersten, and John E. Simmons. *Foundations of Museum Studies: Evolving Systems of Knowledge*. Santa Barbara, CA: Libraries Unlimited, 2014.

Wythe, Deborah, ed. *Museum Archives: An Introduction*. Chicago: Society of American Archivists, 2004.

Organization of Information on the Semantic Web and the Internet

Antoniou, Grigoris, Paul Groth, Frank van Harmelen, and Rinke Hoekstra. *A Semantic Web Primer*. 3rd ed. Cambridge, MA: MIT Press, 2012.

Berners-Lee, Tim, James Hendler, and Ora Lassila. "The Semantic Web." *Scientific American* 284, no. 5 (May 2001): 34–38, 40–43.

DeWeese, Keith P., and Dan Segal. *Libraries and the Semantic Web*. San Rafael, CA: Morgan & Claypool, 2015.

Guerrini, Mauro, and Tiziana Possemato. "Linked Data: A New Alphabet for the Semantic Web." *Italian Journal of Library, Archives, and Information Science* 4, no. 1 (2013). https://www.jlis.it/article/view/6305.

Heath, Tom, and Christian Bizer. *Linked Data: Evolving the Web into a Global Data Space*. San Rafael, CA: Morgan & Claypool, 2011.

Van Hooland, Seth, and Ruben Verborgh. *Linked Data for Libraries, Archives and Museums: How to Clean, Link and Publish Your Metadata*. Chicago: American Library Association, 2014.

Organization of Information in Digital Collections

Calhoun, Karen. *Exploring Digital Libraries: Foundations, Practice, Prospects*. Chicago: Neal-Schuman, 2014.

Fox, Edward A., ed. *Digital Library Technologies: Complex Objects, Annotation, Ontologies, Classification, Extraction, and Security*. San Rafael, CA: Morgan & Claypool, 2014.

Jones, Catherine. *Institutional Repositories: Content and Culture in an Open Access Environment*. Oxford: Chandos, 2007.

Landis, William E., and Robin L. Chandler, eds. *Archives and the Digital Library*. Binghamton, NY: Haworth Information, 2006.

Miller, Steven J. *Metadata for Digital Collections: A How-to-Do-It Manual*. New York: Neal-Schuman, 2011.

Weiss, Andrew. *Using Massive Digital Libraries: A LITA Guide*. Chicago: American Library Association, 2014.

Information Architecture

Dillon, Andrew, and Don Turnbull. "Information Architecture." In *Encyclopedia of Library and Information Sciences*. 3rd ed., edited by Marcia J. Bates and Mary Niles Maack, 2361–68. New York: Taylor & Francis, 2009.

Ding, Wei, and Xia Ling. *Information Architecture: The Design and Integration of Information Spaces*. San Rafael, CA: Morgan & Claypool, 2010.

Robins, David. "Information Architecture, Organizations, and Records Management." *Records and Information Management Report* 17, no. 3 (March 2001): 1–14.

Rosenfeld, Louis, Peter Morville, and Jorge Arango. *Information Architecture: For the Web and Beyond.* 4th ed. Sebastopol, CA: O'Reilly, 2015.

Indexing and Abstracting

Cleveland, Donald B., and Ana D. Cleveland. *Introduction to Indexing and Abstracting.* 4th ed. Santa Barbara, CA: Libraries Unlimited, 2013.

Hedden, Heather. *The Accidental Taxonomist.* Medford, NJ: Information Today, 2010.

Keyser, Pierre de. *Indexing: From Thesauri to the Semantic Web*. Oxford: Chandos, 2012.

Neal, Diane Rasmussen, ed. *Indexing and Retrieval of Non-text Information*. Berlin: De Gruyter Saur, 2012.

O'Connor, Brian C., Jodi Kearns, and Richard L. Anderson. *Doing Things with Information: Beyond Indexing and Abstracting*. Westport, CT: Libraries Unlimited, 2008.

Records Management

Bennick, Ann, and Judy Vasek Sitton. *Managing Active Business Records*. Overland Park, KS: ARMA International, 2014.

Brown, Caroline, ed. *Archives and Recordkeeping: Theory into Practice*. London: Facet, 2014.

Center, Clark E., Jr. *Records Management*. Washington, DC: Association of Research Libraries, 2008.

Saffady, William. *Managing Electronic Records*. New York: Neal-Schuman, 2009.

Smallwood, Robert F. *Managing Electronic Records: Methods, Best Practices, and Technologies*. Hoboken, NJ: John Wiley & Sons, 2013.

Personal Information Management

Jones, William, and Jaime Teevan, eds. *Personal Information Management*. Seattle: University of Washington Press, 2007.

Whittaker, Steve. "Personal Information Management: From Information Consumption to Curation." *Annual Review of Information Science and Technology*, Volume 41. Medford, NJ: Information Today, 2007.

Knowledge Management

Dalkir, Kimiz. *Knowledge Management in Theory and Practice*. 2nd ed. Cambridge, MA: MIT Press, 2011.

Desouza, Kevin C., and Scott Paquette. *Knowledge Management: An Introduction*. New York: Neal-Schuman, 2011.

McInerney, Claire R., and Michael E. D. Koenig. *Knowledge Management (KM) Processes in Organizations: Theoretical Foundations and Practice*. San Rafael, CA: Morgan & Claypool, 2011.

Wallace, Danny P. *Knowledge Management: Historical and Cross-disciplinary Themes.* Westport, CT: Libraries Unlimited, 2007.

* In the profession, the institution is generally known as *an archives*.
* This section was written by Dr. Kyong Eun Oh, Assistant Professor, School of Library and Information Science, Simmons College.

CHAPTER 2

DEVELOPMENT OF THE ORGANIZATION OF RECORDED INFORMATION IN WESTERN CIVILIZATION[*]

It is often said that you can't tell where you're going until you know where you've been. This chapter looks at where we've been and addresses these questions: How did we arrive at the current state of organization? What basic principles of organization have been developed over the last several centuries? Practices and ways of organizing that we now take for granted were once thought of for the first time by intelligent and serious scholars, just as we are coming up with innovative ideas for today's organization that will be taken for granted in the next decades. The history outlined below is primarily a story of the Western development of organization principles. This is not intended to diminish Eastern activities. However, the focus of this chapter is to contextualize the organization of information as it has developed in the West.

INVENTORIES, BIBLIOGRAPHIES, CATALOGS, AND CODIFICATION

Antiquity

One of the oldest lists of books we know about appears on a Sumerian tablet found at Nippur from about 2000 BCE. The tablet records 62 titles, of which 24 are of currently known literary works. We don't know what purpose the list served. Its use may or may not have resembled that of a catalog.[1] However, the fact that the Sumerians were indefatigable writers makes it hard to believe that they had no catalogs. They seem to have kept everything: history books, medical prescriptions, love poems, business invoices, school children's homework assignments, and the first-known letter home from a student who threatens to drop out of school unless his parents fork over more money for a suitable wardrobe.

Through the archaeological discoveries of excavations of ancient civilizations, we know that tablets and other resources were used to inscribe titles of books, but we do not know for what purpose. They might have been ownership tags (e.g., the ones that had the names of the king and queen and a title on each small plaque), or they might have been relics of something like a bibliography or a catalog. There are more remnants of early records from Babylonia than from Egypt, probably due to the fact that Babylonians wrote on clay tablets, while Egyptians wrote on less-enduring papyrus.

Around 1500 BCE, the Hittites evidently saw the need to convey bibliographic information as part of a written work. Their tablets bore colophons that identified the number of the tablet in a series, its title, and often the name of the scribe. A *colophon* is a set of data at the end of a document that gives varying kinds of bibliographic data. It might give information usually found on a title page, and, in items after the invention of printing with moveable type, it gives such information as date of printing, printer, typeface used, and so on.

Around 650 BCE, Ashurbanipal (sometimes written as Assurbanipal) developed a library in the city of Nineveh. This library comprised nearly 30,000 clay tablets and was arranged by subject in a series of rooms. Identification tags were affixed to jars that contained the tablets. Additionally, lists on walls could be found in each room to detail what works were found in that room. This early example of a library is often discussed because of the great care taken to preserve order and authenticity as well as its early catalog-like activities, despite not having anything beyond these informal management tools.[2]

Two of the great libraries of antiquity were in Pergamum and Alexandria—two active centers of Greek civilization. Later writings have referred to *pinakes* from both libraries. Pinakes is the plural of *pinax,* a word that means "tray" or "dish." It is thought that such trays had slightly raised edges and that wax could be poured in the middle; when hardened, the wax could be written in with a stylus. If this was indeed the medium, it is no wonder that no remnants have survived. Writers have quoted from the pinakes of Alexandria, which was created by Callimachus. The work may have been a catalog, or it may have been a bibliography of Greek literature. Callimachus has been given credit as being the first cataloger of whom we have knowledge.[3]

There are a few generalizations about bibliographic practices of the time that can be drawn from quotations of scholars who quoted from Callimachus's work. For example, a few general categories were considered to be a sufficient subject approach. Callimachus's subjects were epic and other non-dramatic poetry, drama, law, philosophy, history, oratory, medicine, mathematical science, natural science, and miscellanea. A scholar would go to the general subject and then look for the author being sought. The arrangement of *entries* (i.e., items in the list) under the general subject was sometimes classified, sometimes chronological, and sometimes alphabetical, although the Greeks never arrived at a strictly alphabetical arrangement. They sometimes grouped entries by initial letter, but there are no examples of arrangement by any letter past the first one. This probably indicates that their lists were not nearly as long as ours.

Greek civilization seems to have given us the basis for our Western idea of the primacy of the author. This creates the idea that the *main entry*, or the thing that dictates where an item falls in the list, ought to be the author. This kind of entry has not appeared in any work that has survived from early Eastern civilizations. Even today in Asian countries the traditional entry for a book is its title. A Japanese librarian of one author's acquaintance once observed that the principle of author entry goes along with democracy, since it rests upon belief in the importance of the individual.

Little information is available about Roman libraries. From sources that mention them, there is evidence that there was some way of finding a designated book when it was requested. This was probably through shelf arrangement based on *fixed location* (i.e., a set physical space assigned to each item). One story goes that if a nobleman got into an intellectual argument, he would send a servant to the library to retrieve a certain book that would prove his point.

Middle Ages

We know that during the Middle Ages in Europe there were church and monastery libraries. Outside of those enclaves, there was little to no demand for books, and knowledge was not sought in any way that would require the use of catalogs. Through copying by monks, a system was set up by which the monastery became the sole keeper, manufacturer, and finally, list maker of books through many centuries.

One of the earliest listings of the holdings of a medieval library was dated in the eighth century. It was written on the final flyleaf of a book and consisted of a list of brief titles with authors added to some of them. It probably served as an *inventory* (i.e., a complete list of resources owned by an institution or a person) and may have represented the shelf arrangement, although there were no location symbols accompanying the titles. This list was typical of most of the so-called catalogs of the following centuries—the briefest sort of inventories recorded in the most casual places.

From the ninth through the thirteenth centuries, libraries continued to produce lists that seemed to be inventories. One list, which specified that its purpose was for inventory, stated that the library contained 246 volumes. It would be quite unrealistic to expect libraries of this size to feel any need for catalogs. Even after libraries grew to the size of 600 or 700 volumes, lists were still inventories. Occasionally such a list would use author entries, but in no discernible order. A few lists gave works contained in each volume, and the number of volumes to a work. (Books into which works were copied were often bound blank pages. Works were copied into them in the order in which the scribe picked them up. Several works could be copied into one bound blank volume, but it might take several bound volumes to copy a very long work.)

The library established in the sixth century by Cassiodorus is the earliest example of a subject arrangement.[4] Subjects during the medieval period were broadly defined, some using only two categories: *biblical* and *humanistic*. By the mid-fourteenth century, however, more sophisticated classification systems began to emerge. One such example is the 1338 catalog for the Sorbonne, which included 59 different classes.[5] At least one list from the thirteenth century added some unusual descriptions in designating books variously as *useless*, *legible*, *old*, and *good*, but we do not know whether they were used as an aid in identification for inventory or to help the reader by pointing out which books could be more or less easily read.

Toward the end of the thirteenth century, someone whose identity is not known started a project that might be considered a milestone in the history of catalogs. This was the compilation of the *Registrum*

Librorum Angliae, a union list of holdings of English monastery libraries in which, in a quite modern way, each library was assigned a number for coding purposes. The *Registrum* was never finished. There is evidence of later attempts to compile continuations of it, although no finished version has survived.

European Renaissance

The fourteenth century brought some improvements, and a few lists of this period might be called *shelflists* (i.e., lists of resources ordered as they are arranged on the bookshelves). The outstanding list of the fourteenth century, compiled by John Whytefield, is for the collection at St. Martin's Priory at Dover, dated 1389. In fact, it may be the first of the lists that could be justly designated a catalog. It is divided into three sections. The first is a listing by a number that represents the fixed location of individual volumes. The second section of the catalog is also arranged by these numbers but gives the contents of each volume as well, including pagination and the opening words for each work. The third part, however, is an innovation in the development of cataloging: a list of analytical entries and an alphabetical listing. (An *analytical entry* is an entry made for each work in a volume, as opposed to making only one entry for the entire volume.) The entries consist of the usual medieval type—some under author, others under title followed by author, with others beginning with words such as *book, part,* or *codex*. There is no obvious importance attached to the entry word.

As monasticism declined, educational institutions began to flourish. At the outset, the best library collections were property of individual faculty members, and students would access necessary resources from their tutors.[6] Centralized college libraries began to be available in the fourteenth century but did not bring any innovations to the development of bibliographic control. The earliest lists from college libraries retain the primitive inventory style of the preceding centuries. As with the monastic libraries before them, college library collections were small and, therefore, required fewer mechanisms for bibliographic control. In fact, it was not unusual for a college library at that time to have only 100 books.

The main new practice from the fifteenth century was the use of *cross-references* (or, simply, *references*). A reference is a pointer from one part of the catalog where someone may logically search to another place where the information is actually found. In one catalog, the references were not separate entries but were appended to a sort of contents note pointing out in what other place in the library a certain item might be found (e.g., "which seek in the 96th volume of theology"). In the catalog of St. Augustine's Abbey, Canterbury, though, references reached the status of separate entries.[7] A typical example of the kind of reference employed is

> The Meditations of Bernard, not here because it is above in the Bible [which was given by] W. Wylmynton.

In the middle of the fifteenth century came an event that challenged everything about bibliographic control—the invention of moveable type printing by Johannes Gutenberg in 1455. Suddenly, instead of unique manuscript copies of works, there were identical duplicates of many works. The invention of the printing press was a watershed moment in Western history that reverberated in all cultural, economic, and political spheres. It should be no surprise that it had a significant impact on the bibliographic universe.

The invention of the printing press ushered in the need for bibliographic control previously unnecessary due to small collection sizes. This need created a new area of expertise, *bibliography*: the study of books as physical and cultural objects, which often results in creating extensive lists. Toward the close of the fifteenth century, German abbot and bibliographer Johannes Trithemius (also known as Johann Tritheim) stands out as having taken an important step in the development of bibliographic control. He not only compiled a bibliography in chronological order, which was unusual enough for his time, but he also appended to this an alphabetical author index. It is difficult to understand why such a simple and useful device had not always been used; yet it took centuries of compiling book lists to reach this degree of accomplishment.

From Inventories to Finding Lists to Collocating Devices

Bibliographers in the sixteenth century continued to take the lead in making improvements to bibliographic control devices. One of these, Conrad Gessner (also known as Konrad Gesner), published an author bibliography in 1545 and a subject index in 1548, and in the process set a new standard of excellence. He continued to use forenames (i.e., first names or given names) of authors for entry words, according to the tradition of the time, but he recognized the possible inconvenience of this practice and so he prefixed to his bibliography an alphabetical list of authors in which the names were inverted. In addition, his main listing included references from variant spellings of names to the accepted entry form (e.g., "Thobias, see Tobias"). Gessner included in his work (titled the *Pandectarum*) the suggestion that libraries use copies of his bibliographies as their catalogs by inserting call numbers beside entries that represented their holdings, thus providing themselves with both an author and a subject catalog.

The end of the 16th century saw even greater advances in bibliography, and some of the conventions that we now take for granted were introduced. For example, in 1595 Andrew Maunsell, an English bookseller, compiled his *Catalogue of English Printed Books* and in the preface stated his rules for entry. He advocated the entry of personal names under *surnames* (i.e., family names or last names) rather than forenames. He set up the principle of uniform entry for the Bible, copies of which, prior to Maunsell's collocating them, had been entered under whatever the title page said (*Holy Bible*; *The Word of God*; *Bible*; or another title). He insisted that one should be able to find a book under three types of entries—the author's surname, the subject, and the translator. These were radical and sudden advances in the development of bibliographic control.

By the beginning of the seventeenth century, catalogs were beginning to be looked upon as finding lists rather than inventories. Early in the century, Sir Thomas Bodley offered to build up the Oxford University Library, which had been dissolved some 50 years before when, in 1550, the dean of Oxford's Christ Church college, "hoping to purge the English church of all traces of Catholicism including 'superstitious books and images,' removed all the library's books—some to be burnt."[8] Bodley took a great interest in the catalog because he expected that it would be useful in his acquisitions program; he wanted the catalog to tell him if the library already owned a work. He insisted upon a classified catalog with an alphabetical author index arranged by surname, and he also wanted analytical entries.

In 1697 Frederik Rostgaard published a discourse on cataloging in which he called for subject arrangement subdivided at once chronologically and by size of volume. For the preceding century, size of volume had been a traditional way of dividing catalogs. Rostgaard proposed a printed catalog, with the spread of two facing pages divided into four parallel columns, each column to contain books of a certain size (folios, quartos, octavos, and duodecimos), arranged so that books of various sizes that had been published on a certain subject within the same year would come directly opposite each other in parallel columns. He recommended an alphabetical index of subjects and authors to be placed at the end of the catalog, with authors entered by surname. The word order of titles as found on the title page was to be preserved. His final suggestion was that his rules not be followed when it seemed best to arrange things differently!

As the eighteenth century began, bibliographic control seemed to have hit a plateau that did not change for most of the century. Catalogs were sometimes classified and sometimes alphabetical and usually divided according to the size of books; indexes were considered useful, though by no means necessary; authors were now always entered under surname and were often arranged chronologically; the wording of the title page had assumed a certain degree of prestige and was now being transcribed literally and without being paraphrased; imprints were included; bound-with notes were used; cross-references were quite common; and some analytical entries were used in most catalogs.

The French Revolution provided the impetus for the creation of a new kind of catalog. In 1791 the new French government sent out instructions for cataloging the collections of the libraries that had been confiscated throughout the country. These collections were to be sold to help cover the government's expenses or to be kept to form a national library, open to the public, to engender the diffusion of knowledge, believed to be necessary in a democratic government. To fulfill this program, they created the first instance of a national *cataloging code* (i.e., a set of guidelines or instructions on how to catalog). The need for the code was twofold: first, the inexperience of the catalogers required some kind of instruction; and second, and more importantly, uniformity would allow identification of duplicate items, facilitating decisions about the disposition of particular copies.[9] The instructions included the use of playing cards for recording their catalog information:

> That there might be enough blank space for writing, it was recommended, in case of very long titles, that the playing cards chosen should be those with the smallest number of pips, as the ace, deuce, etc., and with as plain backs as possible, so that, if the space for the title was insufficient on the face of the card, the back might be used to complete it.[10]

This is possibly the first appearance in history of a *card catalog*. It was introduced, not because someone thought it would be a convenient form, but because, with wartime shortages, it was a practical way of using available materials. Confiscated playing cards were to be used for the purpose. Playing cards of the time were blank on the back, rather than having pictures.[11] They were also larger than today's cards.

It is notable that even though the card catalog of the French Revolution is thought to be the first card catalog recorded in history,[12] and it provided the first national instructions on how to create card entries, the catalog is itself unremarkable. The purpose of the code was not to provide a catalog of a library for access purposes but to serve an administrative function for the creation of a union catalog that would assist in the appropriate confiscation of unique resources. As stated in the instructions,

> It is necessary to set up catalogs having no other object than of procuring an exact knowledge of all the books, printed as well as manuscripts, that exist in those libraries of each department that make up part of the "biens nationaux."[13] [*Biens nationaux* refers to goods confiscated from the church, the monarchy, and other elites during the Revolution.]

That this catalog was not meant for public consumption surely affected the construction of the catalog itself. The rules themselves contain instructions on the physical process of cataloging as well as on the intellectual content needed. There was no theory or philosophizing in this code. The title page was to be transcribed on the card and the author's surname underlined for the filing word. If there was no author, a keyword in the title was to be underlined. A *collation* was added that was to include number of volumes, size, a statement of illustration, the material of which the book was made, the kind of type, any missing pages, and a description of the binding if it was outstanding in any way. (This elaborate collation was partly for the purpose of identifying valuable books that the government might offer for sale in order to increase government revenue.) After the cards were filled in and put in order by the underlined filing word, they were to be strung together by running a needle and thread through the lower left corners to keep them in order.[14]

Despite the lack of intention, the French Revolution card catalog established a number of procedures that have continued to the present. The resulting cataloging code was the first systematic national effort to standardize cataloging, and it is notable for its use of cards in the production of a catalog. Although it would be another 100 years before card catalogs became the standard for libraries in the Anglo-American world, this code, coming at the end of the eighteenth century, provides a good stepping-stone to the extensive cataloging developments of the nineteenth century.

Period of Codification

The nineteenth century brought a period of much debate over the relative virtues of classified and dictionary catalogs, not only among librarians, but also among readers and scholars in general and even in reports to the House of Commons of Great Britain. Feelings ran very high on the subject, and rather emotional arguments were made, "from the statement that classified catalogs and indexes were not needed because living librarians were better than subject catalogs, to the opinion that any intelligent man who was sufficiently interested in a subject to want to consult material on it could just as well use author entries as subject, for he would, of course, know the names of all the authors who had written in his field."[15]

What was needed was a person who could persuade others of the value of cataloging and subject analysis. That person turned out to be Anthony Panizzi, a lawyer and Italian political refugee who was appointed assistant librarian at the British Museum in 1831. When he was appointed Keeper of the Printed Books in 1837, there was much objection. One historian, Dorothy May Norris, states that it was because "firstly, Panizzi was an Italian by birth, and it was felt that only an Englishman should be in charge of one of our national institutions; secondly, it was said that Panizzi had been seen in the streets of London selling white mice."[16] No further explanation is given. Louis Fagan, another writer, gives this account: "Meetings were held against the '*Foreigner;*' and one of the speakers made an open statement that Panizzi had been seen in the streets of London selling *white mice:* had it been a few years later, possibly the distinctive title of organ-grinder would have been added."[17]

In 1836 a committee of the House of Commons was charged with inquiry into the management and affairs of the British Museum. One of the "affairs" was the state of catalogs and cataloging. During hearings on the topic, witnesses came to testify for and against the catalogs. Many of the witnesses became quite vehement about one or another sort of entry. Panizzi was able again and again to persuade the committee members to accept his views. Panizzi wrote his views into a cataloging code titled, "Rules for the Compilation of a Catalogue" (more commonly known as the "91 Rules"). It appeared in the first and only volume of *The Catalogue of Printed Books in the British Museum*.[18] This code gained official approval in 1839, although Panizzi had to give up his concept of corporate authorship in order to get approval. (This concept was opposed by those who believed that only individual persons are capable of writing.)

It is generally accepted that the catalog code of the British Museum, published in 1841, begins the current period of cataloging. The legacy of the "91 Rules," however, is not as simple as merely being the first formal English-language catalog code. While Panizzi managed to create only the first volume of his planned catalog due to political infighting, the rules themselves provided a foundation upon which later codes could rest. Panizzi's code shows that we had, at last, arrived at modern cataloging, because he tried to deal with many of the same problems we are still arguing about today.

In the United States, cataloging enjoyed the same growing pains as its European counterpart. Prior to the 19th century, American cataloging had been generally a century behind European cataloging. For example, the earliest catalogs printed at Harvard were divided by size and then alphabetized according to author, title, or keyword, with shelf locations and very brief records, but no subject access.[19] Preceding the introduction of mechanized card production in 1901 by the Library of Congress (LC), much of the cataloging performed in the United States resulted in book catalogs, both print and manuscript.[20] Cataloging was one area under scrutiny by the leadership of the nascent profession of librarianship. Charles Coffin Jewett is one of the leaders who placed this activity at the center of his work.

Jewett was a librarian during a period when printed catalogs were considered premier. He was trained in the practice of cataloging at the beginning of his career. Throughout his career, he was considered to be an expert in the practice of constructing and evaluating library catalogs. He enjoyed success at the Brown University Library, and then was appointed as librarian for the newly established Smithsonian Institution. Later in his career, he had a significant impact on the Boston Public Library as superintendent.

In 1850 Jewett published a code for the catalog of the Smithsonian Institution. With this code Americans began to have influence in cataloging. Jewett acknowledged his debt to Panizzi and varied in only a few instances from the instructions in the "91 Rules." Jewett is given credit for extending the principle of *corporate authorship* further than Panizzi had. Research now shows that Jewett copied his rule from Panizzi's original draft, which had a rule for corporate bodies as authors but which did not appear in the "91 Rules," because Panizzi was forced by the British Museum trustees to drop his rules for entry under corporate author.[21] So what Jewett actually did was to bring the concept of corporate authorship to public attention.

Jewett's philosophy of the purpose of a code was this: "*Uniformity* is, then, imperative; but, among many laborers, can only be secured by the adherence of all to rules embracing, as far as possible, the minutest details of the work."[22] In light of this philosophy, it is interesting to observe that the second edition of Jewett's rulebook contains only 39 rules on pages 29–64 with pages 67–90 devoted to examples. The title of Jewett's book, *On the Construction of Catalogues of Libraries, and Their Publication by Means of Separate, Stereotyped Titles,* makes it clear that Jewett was addressing more than just rules for cataloging. The first part of the volume is devoted to his plan to create a stereotyped plate of each title cataloged, to store the plates, and to use them over and over for production of printed catalogs. *Stereotyping* was a method of printing using a metal copy of a typeset image. Jewett planned to have the plates numbered in the order in which they were printed and to keep them in alphabetical order in shallow drawers. The numbers could be used by libraries to indicate which titles were owned by each, and then a catalog could be produced for a library by pulling out the plates for titles owned by that library. Also, a union catalog could be produced from the plates. Jewett's plan was never implemented. Apparently, he was technologically a bit ahead of his time.[23] To say that Jewett influenced American librarianship—from his contributions to cataloging, his vision of a national library (which he hoped would be the Smithsonian), and his leadership in the burgeoning library profession, particularly at the 1853 Librarian's Convention (often credited as a precursor to the American Library Association)—would be an understatement. Even his contemporaries, without the benefits of hindsight, recognized Jewett as a visionary in cataloging and a leader in librarianship. While Jewett made significant contributions to American librarianship and cataloging specifically, his contributions are overshadowed by those of Charles Ammi Cutter and Melvil Dewey (discussed below).

Cutter, as librarian of the Boston Athenaeum, created his masterwork, *Rules for a Dictionary Catalog*, which ran to four editions and can be credited as the foundation of American cataloging. His set of cataloging rules still informs our cataloging standards today. Cutter, along with Dewey and others, was instrumental in the establishment of the American Library Association just over two decades after the 1853 Librarian's Convention. Like Jewett before him, Cutter was considered a leader in the profession. He also created the *Expansive Classification* system and was the genius behind book numbers, still in use today and known as *Cutter numbers*. His legacy, though, rests squarely in his articulation of the objectives of the catalog,[24] in addition to his innumerable contributions to cataloging, the profession, and the larger library mission.

When Cutter published his *Rules* in 1876, he strengthened the concept that catalogs not only should point the way to an individual publication but also should assemble and organize literary units.[25] That is, catalogs should be collocating devices. This was not an entirely new principle: Maunsell had used the heading **Bible**;

Panizzi had strengthened it by introducing corporate and government entries; and Jewett had given further support by use of real names rather than pseudonyms. But it was Cutter who actually stated it as a formal principle.

Cutter was also the first to make rules for subject headings as a way to gain subject access to materials through the catalog. And he was the last to incorporate into one set of rules instructions for the description of items, guidelines for subject headings, and rules for filing entries. At the end of the nineteenth century each of these areas (i.e., description, subject headings, and filing) took on lives of their own and followed separate paths of development. We will now follow the first two paths separately. (Filing arrangement is discussed in Appendix C: Arrangement of Metadata Displays.)

THE TWENTIETH CENTURY

Description

In the area of description, the twentieth century was an era of codes. The British and the Americans collaborated on a code published in 1908. Its importance lies in its being the first *international* cataloging code, and in the extent of its rapid and widespread adoption and use by all types and sizes of libraries in the two countries.

In the 1920s four prominent American librarians helped with writing the Vatican Code, which was published in Italian in 1931. It was quickly accepted by catalogers in many countries as the best and most complete code in existence, but because it was in Italian, most Americans could not use it.

The British and Americans began cooperative work toward a new edition of the 1908 code in the 1930s, but the outbreak of war ended this cooperation. The American Library Association (ALA) proceeded independently in producing its preliminary second edition in 1941. This code was published in two parts: one for entry and heading, and one for description of books. It was widely attacked for its complexity and the proliferation of rules for every circumstance. The most famous attack was launched by Andrew Osborn in his article, "The Crisis in Cataloging."[26] Osborn identified four principal theories of cataloging:

- **Legalistic—**cataloging with a rule for everything, and an authority to settle any question at issue,
- **Perfectionistic—**cataloging performed so well the first time that it is done once and for all,
- **Bibliographic—**cataloging made into a branch of descriptive bibliography with extremely detailed physical descriptions and notes, and
- **Pragmatic—**cataloging according to the needs of particular types of libraries and/or users.

He called for both the cataloging process and the rules governing it to become pragmatic, with simple rules supplemented by cataloger judgment. Osborn's article is one of the classic statements in cataloging theory and certainly one of the historical turning points in code development.

In response to all the criticism, the ALA Division of Cataloging and Classification undertook revision of the first part (entry and heading) of the 1941 code. In 1949 the revision was published as the *A.L.A. Cataloging Rules for Author and Title Entries.* The *Rules for Descriptive Cataloging* published by LC was substituted for the second part (description of books) of the 1941 rules. The ALA portion of the 1949 rules again was criticized as being a continuation of Osborn's "legalistic" characterization. In 1951 the ALA commissioned Seymour Lubetzky to do a critical study of cataloging rules. Lubetzky said cataloging should be done according to principles, and he drafted a code based on principles.[27] Progressives welcomed it, but conservatives began worrying about probable costs of changes.

An International Conference on Cataloguing Principles was held in Paris in 1961, the first major initiative for the International Federation of Library Associations and Institutions (IFLA), at which a draft statement of principles, based on Lubetzky's code, was submitted. The goal of this conference was to agree on principles that would inform cataloging standards internationally. The participants agreed to adopt these principles and to work in their various countries for revised rules that would be in agreement with the accepted principles. These principles, often referred to as the *Paris Principles* (or *IFLA Principles*) are important because, for the first time, there was multinational agreement upon which to base future international developments.

The Americans and the British again cooperated on a new set of cataloging rules, and in 1967 published the *Anglo-American Cataloging Rules* (although it had to be published in separate North American and British versions because of inability to come to agreement on some points). These rules, known as AACR, were based on the Paris Principles, although they deviated in some respects.

In 1974 IFLA issued the *International Standard Bibliographic Description* (ISBD), produced as a means for the international communication of bibliographic information. ISBD's objectives were to make records from different sources interchangeable, to facilitate their interpretation against language barriers, and to facilitate the conversion of such records to machine-readable form. To meet these objectives, ISBD provides a standard set of bibliographic elements with which to describe resources, a standard order for those elements, and a system of punctuation symbols to be used to identify individual elements (this allows for the identification of individual elements when the language of the cataloging data is unfamiliar).

AACR2, the second edition of the *Anglo-American Cataloguing Rules,* was published in 1978. This new edition sought to incorporate ISBD, to bring nonbook materials into the mainstream, to take into account machine processing of bibliographic records, to reconcile the British and American texts, and to conform more closely to the Paris Principles. Four major national libraries (United States, Canada, Great Britain, and Australia) agreed to standard interpretation and implementation of AACR2, and many other countries later adopted it for national use. Revised editions of AACR2 were published in 1988 and 1998. These were mainly cumulations of changes incurred in the process of continuous revision. A significant revision of AACR2 was published in 2002, with updates distributed in 2003, 2004, and 2005. Updates for AACR2 ceased in 2005, as efforts were focused on the creation of a new descriptive cataloging content standard, *RDA: Resource Description & Access,* which was published in 2010. As was the case with AACR2, implementation of the new standard was delayed by several years. In the United States, RDA was not adopted by the national libraries (LC, the National Agricultural Library, and the National Library of Medicine) until 2013.

Subject Access

Verbal Subject Access

For centuries philosophers worked on classifying knowledge; Callimachus, Plato, Aristotle, and Bacon are among the most famous. Librarians tried to adapt these classifications for use with books by assigning letters and/or numbers to the concepts classified by the philosophers. Other than this, there was not much interest in subject access in libraries before Charles Cutter. As already mentioned, Cutter included a section of guidelines for subject headings in his *Rules for a Dictionary Catalog.*[28]

The ALA's *List of Subject Headings for Use in Dictionary Catalogs* was first published in 1895, and the preface stated that it was to be considered an appendix to Cutter's *Rules for a Dictionary Catalog.* It was based on headings found in five major catalogs of the time, including the Boston Athenaeum and the Harvard subject index. A second revised edition, "with an appendix containing hints on subject cataloging and schemes for subheads under countries and other subjects," was published in 1898 with the statement that "further changes are not to be expected for many years."[29] However, new terminology became necessary rapidly, and interleaved and annotated editions became unwieldy. Many librarians asked for the list of subject headings that were appearing on LC cards, and so from 1909 to 1914 the first edition of *Subject Headings Used in the Dictionary Catalogues of the Library of Congress* was published in parts. This list rapidly replaced the ALA list as the distribution of LC cataloging cards spread. The title of LC's list was changed to *Library of Congress Subject Headings* (LCSH) in 1975. LCSH has appeared in multiple editions in its printed version, and from 1988 to 2013 a new edition was published every year (or nearly). The 35th edition was the last print edition of the "the big red books" (as they were sometimes called). Although the print editions are no longer available, the system of subject headings is available electronically in different forms. Many institutions access LCSH through a searchable, interactive, web-based subscription service, *Classification Web*.[30] This service, however, may be cost prohibitive for smaller libraries. Others access LCSH through the LC subject authority file that is freely available at LC's website or through a subscription to a bibliographic network, such as OCLC Online Computer Library Center. Starting in 2014, LC began providing free access to LCSH through print-ready downloadable PDF files. Each year, a new edition is published on the LC website.[31] LC has adopted a process of continuous revision for its subject heading list; new and revised headings are approved monthly.

The *Sears List of Subject Headings (Sears)* was first published in 1923 as *List of Subject Headings for Small Libraries.* It was prepared by Minnie Earl Sears in response to demands for a list of subject headings that was more suitable to the needs of the small library than the ALA or the LC lists. Recognizing the need for uniformity, Sears followed the form of the LC subject headings, although she eliminated the more detailed headings and simplified some terminology. *Sears* continues to be published in print for small libraries. The online version, available by subscription through EBSCOhost, is updated annually.[32]

Classification

Meanwhile, classification had developed at LC from arrangement by size in the early 1800s to arrangement by the 18 broad categories of the Bacon-d'Alembert system in 1814 when LC was burned by the British during the War of 1812. To reestablish the library's collection, Thomas Jefferson sold Congress his library, which was classified using 44 main classes based on his interpretation of the Bacon-d'Alembert system. By the end of the century it was clear that the rapidly growing collection needed more detail in classification.

Melvil Dewey, in 1876, issued anonymously the first edition of his classification. He divided all knowledge into 10 main classes, with each of those divided again into 10 divisions, and each of those divided into 10 sections—giving 1,000 categories into which books could be classified. Like its predecessors, it was enumerative in that it listed specific categories one by one. In later editions he added decimals so that the 1,000 categories could be divided into 10,000, then 100,000, and so on. He also introduced the first hints of *number building* (or *faceting*)[33] when he made tables for geographic areas and for forms of material. Notations from these tables (today referred to as *subdivisions*) could then be appended to the base number to show a certain subject as being relevant to a particular geographic area or being presented in a particular form.

By the end of the nineteenth century, LC had decided it had outgrown Jefferson's classification. The *Dewey Decimal Classification* (DDC) was in its fifth edition, and Cutter had begun his own enumerative classification scheme, called the *Expansive Classification* (EC). This scheme began with letters of the alphabet, expanded with second letters, and then expanded further with numbers for different facets. LC representatives talked with Dewey to try to convince him to allow them to adapt his scheme but ran afoul of his intransigence. They did not adopt Cutter's classification directly but created their own scheme based upon his main class structure. This was the origin of the *Library of Congress Classification* (LCC).

The *Universal Decimal Classification* (UDC) was developed in 1895 by two Belgian lawyers, Paul Otlet and Henri La Fontaine. UDC was based on the DDC (then in its fifth edition) but was, with Dewey's permission, expanded by the addition of detailed subdivisions and the use of symbols to indicate complex subjects. Why Otlet and La Fontaine were able to get Dewey's permission when LC was not is unclear; but it may have been because UDC was not at the outset intended as a library classification but rather as a means to organize documents (see more about this below). UDC expanded Dewey's standard subdivisions to about a dozen generally applicable auxiliaries, which could be joined together as needed in the form we now call *faceting*.

During the twentieth century, faceted classification became of interest because it allows the classifier to express all aspects of an interdisciplinary subject in the same classification notation. This type of classification is also referred to as *analytico-synthetic classification.* The term *facet* was first used with this meaning by S. R. Ranganathan in the early 1930s in his *Colon Classification* (CC), named for its use of the colon punctuation mark as a major facet indicator. As mentioned earlier, Dewey had already provided for some number building, and LCC even included a few such facets (although developed only class by class and not applicable through the entire scheme). Ranganathan introduced the fully faceted approach by means of classification notations constructed entirely from individual facets in a prescribed sequence from the most specific to the most general. The work of Ranganathan inspired many librarians and information scientists in the second half of the twentieth century, most notably a group of classificationists in Great Britain collectively known as the Classification Research Group (CRG). The CRG's work led to the revision of the *Bliss Bibliographic Classification,* the creation of the *Preserved-Context Indexing System* (PRECIS), and the development of several faceted classifications used in particular disciplines and industries.

Most classification systems have been used, beginning in the late 1800s, as the basis upon which to create call numbers to be used for the physical arrangement of books and other resources on shelves in stacks. Arrangement of physical information resources is discussed in Appendix B.

Special Materials

Archives

U.S. developments in bibliographic control of special materials (i.e., archives, museums) have nearly all come since the turn of the twentieth century. European archival practice stemmed from working with public archives (e.g., land grants, laws). Archival materials were kept because of legal and other administrative value. The concept of *provenance* emerged in France about 1840, and the concept of *original order* came from the Prussians shortly thereafter.[34]

The first articulation of principles for archives was published in 1898 by Dutchmen Samuel Muller, Johan Feith, and Robert Fruin[35] and is typically regarded as the foundation of archival theory. This manual articulates 100 rules that would assist with "uniformity in the handling of inventories both in essentials and in details."[36] The concepts articulated in the manual include the definition of *archives*, the importance of arrangement (including the principle of *respect des fonds*), the organic nature of archival collections, the role of description and its relationship to arrangement, and terminological definitions. All of these concepts appear in later codified international descriptive standards. The manual was quickly adapted by other European countries, first being translated into German, then Italian and French. Bulgarian, Russian, and Estonian editions appeared in the second decade of the twentieth century. It was finally published in English in 1940.[37] As Marjorie Rabe Barritt articulates in her account of its introduction to the United States, "the spread of ideas and theories can be subtle; they often do not wait for translation to begin to effect change."[38] The principles and concepts published in the *Manual* gained traction in the United States through the endorsement of individuals willing to argue for their relevance. American archival practice, before the distribution of the English-language *Manual*, was influenced more by historians and librarians than it was grounded in archival theory. The establishment of the Society of American Archivists in 1936, and the subsequent publication of the *Manual* in English in 1940, would initiate a half-century search for standards and codification.[39]

As in libraries, the first archival catalogs were lists and then inventories. In early archival practice in the United States, material was collected for its artifactual value. It was often cataloged at the item level without any concern for provenance or original order. Collections were sometimes formed based on similar topics. Thus, the context in which the archival record was originally created was lost. This practice lasted in the United States through the 1930s, when the European ideas articulated in the *Manual* began to influence U.S. practice. Early cataloging codes in the United States (e.g., Cutter, AACR) dealt with cataloging manuscripts, but at the item level. In the early 1980s, Steve Hensen, while working for LC, was asked to construct a cataloging code for archival materials. His work resulted in publication of *Archives, Personal Papers, and Manuscripts* (later becoming known as APPM), which brought library and archival traditions together.[40] APPM was based on ISBD and AACR2 but also included archival principles. A second edition of APPM was published in 1989 and served as the content standard for archival description for the next 15 years.[41] APPM provided guidelines for the content of MAchine-Readable Cataloging (MARC) records for the description of archival materials. A special MARC format was developed (called MARC-AMC). It has now been incorporated with the other special MARC formats in the process of "format integration." APPM, though, only provided guidance for MARC records and did not adequately address the needs of finding aids, the primary descriptive tool created by archivists. To that end, *Describing Archives: A Content Standard* (DACS), the descriptive standard used in the creation of finding aids and MARC records for archival collections in the United States, was first published in 2004 and superseded APPM. A second edition of DACS was released in 2013.

Museums (Art and Object Collections)

As in libraries and archives, museums began first with lists and later expanded to include inventories. *Museum documentation* or *museum registration* is terminology that often has been applied to cataloging of art and artifacts. It is a system that provides an indispensable record of information associated with objects for research. Museums are now joining the library and archives communities in codification of descriptive practice. The Internet has spurred this action because of the demand from researchers for access to pictures and textual descriptions of artifacts and art. The visual resource community has developed standard means for creating surrogate records (e.g., *Cataloging Cultural Objects* [CCO]). Libraries with these kinds of collections may use *RDA*, which can be used to describe any type of information object, or they may still use AACR2's chapters for cataloging realia and graphics, and then enter these descriptions into a MARC record.

Subject Access to Special Materials

Subject access to special materials follows the needs of the user communities. For verbal subject access, archivists use LCSH and the *Art & Architecture Thesaurus* (AAT). The museum community has developed specialized lists and thesauri (often called *lexicons*). One example is Robert G. Chenhall's *Nomenclature*, originally published in 1978 and currently in version 4.0. It is a system for classifying human-made objects, which has been adopted by thousands of museums and historical societies in the United States and Canada.[42]

The American Association for State and Local History supports a user community for this resource. Museums also use AAT. The collective nature of archival description does not lend itself to the use of classification. Some collections of art and artifacts are classified, however. Some natural history museums classify specimens of organisms that can be identified to the genus–species level. Two other examples include the American Museum of Natural History, which uses Romer's classification of vertebrae, and art collections with Christian iconographic themes that use *Iconclass,* developed in the Netherlands.

Mechanization of Bibliography

Automation first entered the information organization picture in the 1870s when the typewriter was introduced into libraries. Typewriters were highly controversial at first because they were so noisy. Many libraries were one-room affairs with the cataloger sitting at a table in the back. Patrons, who had been used to the quiet of the cataloger creating handwritten cards, found the clacking of the typewriter annoying. Some librarians objected, too, because typed cards were not considered aesthetically pleasing when compared to cards written in *library hand*, a method of writing in which the letters were carefully formed to be completely readable (see Figure 3.3 on page 108).

The Documentation Movement

The trend to mechanize bibliography began with the Documentation Movement in Europe in the 1890s. The nineteenth century brought the development of professional organizations and the growth of scientific research, both of which created a dramatic increase in the number of published journals. Otlet and La Fontaine spearheaded a movement for bibliographic control in libraries to go beyond books to provide access to parts of books, articles in journals, research documents, brochures, catalogs, patents, government records, archives, photographs, and newspapers. The goal of the Documentation Movement was to capture, record, and provide access to all information in all formats for the improvement of science.

A conference organized by Otlet and La Fontaine was held in Brussels in 1895, with a focus on the creation of Universal Bibliographic Control (UBC).[43] The magnitude of the undertaking necessitated an entirely new body of techniques different from conventional library practice for organization, subject analysis, bibliographic description, and annotation. This quest for new techniques of bibliographic control naturally led to a search for new technology. The concept of documentation was transported to the United States in the 1930s, and in 1937 the American Documentation Institute was formed. In 1938 the International Federation for Documentation was established and was devoted almost exclusively to the promotion of *Universal Decimal Classification* (UDC). From its inception, UDC was not intended as a library classification but rather as a means to organize and analyze *documents,* a term then defined to include such things as journal articles, scientific papers, patents, and the like, but not books or journals.

An important technological advance for the documentation field was the development of microphotography by Eastman Kodak in 1928. This was seen as a means of collecting, storing, and accessing vast quantities of information, and it was predicted that microfilm would supplant the conventional book.

World War II had an impact on the mechanization of bibliographic control in two ways. First, it created an immediate scientific information explosion, as the U.S. government's imperative was to "get the bomb first." Scientific research was conducted rapidly and in secrecy, with a critical need for immediate dissemination of research results from lab to lab. This heightened the government's awareness of the need for bibliographic control and brought with it government funding to develop a mechanized process. Second, as the outcome of the war shifted in favor of the Allies, huge quantities of German scientific literature were confiscated. This material needed immediate document control in order to be useful. Microfilms were made and distributed through a committee attached to the Office of Strategic Service (OSS; a wartime intelligence agency that is the predecessor of the Central Intelligence Agency). The distribution was under the direction of Frederick Kilgour, who later started OCLC. At the same time, the Central Information Division of the OSS was working on the subject analysis of documents using IBM punched card equipment. The war itself was the impetus for many technological advances; some of these advances could now be applied to organizing and accessing the hoards of scientific and technical literature that also were an outgrowth of the war.

In 1945 Vannevar Bush opened the way for a new era in documentation and information science with his article "As We May Think." Bush developed the idea of *memex,* a "device in which an individual stores all . . . books, records and communications, and which is mechanized so that it may be consulted with exceeding speed and flexibility."[44] Bush's memex is, in some ways, similar to Otlet's vision of the *televised book* in the *radiated library*; today, both are seen as precursors of or inspirations for the Internet.[45] Using the medium of microfilm (in 1945, remember), Bush described in detail what has become the hypertext-based scholar's workstation of today. Memex was based on the concept of *associative indexing*, similar to the human thought process, where items are linked together and any item can immediately lead to the access of other related information. Bush even predicted new forms of encyclopedias where information could be coded and connected to pertinent articles. A man of vision, Bush believed that science should implement the ways in which we produce, store, and consult the record of the human race.

The 1950s and 1960s saw many and varied attempts at mechanization using the then current technology. Calvin Mooers coined the phrase *information retrieval* in 1950. He and other information scientists of the day, such as Ralph Shaw and Mortimer Taube, worked on developments such as the *Rapid Selector*, designed to provide subject access to microfilm by a method that used holes punched in the sides of the film. Another technique used a knitting-type needle to access subjects on punched cards. Taube was especially concerned with the linguistic problems of documentary analysis and retrieval.

In 1957 *Sputnik,* the first artificial satellite launched into space from Earth, pushed information needs to the forefront of the scientific community once again. There was increased interest in improving access to recorded knowledge in both the government (the National Science Foundation, the National Library of Medicine, LC, etc.) and the private sector (IBM, General Electric, Kodak, RAND Corp., etc.). The field of documentation became the field of information science with a great deal of money made available for research and development. The 1960s saw a period of tremendous technological advances in communication and information processing. The computer became established as the means of storing massive amounts of data and providing high-speed access. In 1968 the American Documentation Institute changed its name to the American Society for Information Science (ASIS) and changed its name again in 2000 to American Society for Information Science & Technology (ASIS&T). In 2013 it changed its name yet again to Association for Information Science and Technology (also ASIS&T) to reflect its growing international membership. By the late 1960s the use of machines for the retrieval of information was solidly entrenched in the information science community—a community that had developed quite separately from libraries until this point.

Library Automation

In the late 1960s two developments changed the face of information organization forever. At LC, Henriette Avram engineered the creation of the MARC format, enabling the machine readability of bibliographic records. Avram was able to draw on early experiments to create machine-readable entries. A pilot project was embarked on in 1965 under the Office of Information Systems Specialists. Avram was a systems analyst who had recently joined LC. She collaborated with Kay Guiles, a descriptive cataloger, and Ruth Freitag, a reference librarian, to construct an "input format that is reasonable to produce, function codes to explicitly define those elements that cannot be recognized by machine today, and programming rules to define those elements implicitly defined."[46] The work on the development of MARC in the 1960s and early 1970s resulted not only in machine-readable bibliographic data but also in records that facilitated data exchange and opened the door to what has now been nearly 50 years of expansion through automation and data sharing.

In Ohio, the Ohio College Association sought to promote library cooperation in order to maximize resource availability and minimize costs. They brought in Ralph Parker of the University of Missouri and Frederick Kilgour of Yale University to envision a new approach: a cataloging system and union catalog that leveraged emerging technologies. Kilgour went on to become the first executive director of the Ohio College Library Center (OCLC) and to oversee a startling period of growth, starting with six staff members after its first two years and expanding to over 200 staff members after its first decade. Kilgour continued to serve as its director until 1980. Over that period, the online system of shared cataloging became fully operational and extended its membership to libraries outside of Ohio. In 1981 OCLC changed its name to OCLC Online Computer Library Center, Inc. The following three decades saw numerous projects and partnerships that underscored the basic mission of the organization, including the incorporation of technological advances when they occurred.[47] OCLC continues to expand its offerings to the larger library community. It currently offers management services, metadata services, discovery services, and resource sharing services and maintains the most substantial bibliographic network, WorldCat.[48]

With the development of the MARC format, OCLC was able to provide cataloging information via cable and terminal to all its member libraries, which in turn were able to put their original cataloging online for the use of all other members. In 1977 another major network came into being particularly to serve research libraries—the Research Libraries Information Network (RLIN), which in 2006 was absorbed into OCLC.

Other bibliographic networks are operational today (e.g., SkyRiver, Auto-Graphics), but they would not be possible without the pioneering work of Avram, Kilgour, LC, and OCLC.

CONCLUSION

We see here the coming together of the information science track and the library science track, which had previously developed separately. Some see a wide gulf between information science and library science, but both are interested in and working on the organization of information, as we see in the chapters that follow. And the distinctions are becoming less and less obvious as our technology and information infrastructures grow relentlessly intertwined.

Some Important Terms, Historical Figures, Events, and Other Entities in This Chapter

Analytical Entry
Analytico-synthetic Classification
Ashurbanipal
Avram, Henriette
Bibliography
Bodley, Thomas, Sir
Bush, Vannevar
Callimachus
Card Catalog
Cassiodorus
Catalog
Cataloging Code
Collation
Colophon
Cross-reference
Cutter, Charles Ammi
Dewey, Melvil
Entry
Faceting
Feith, Johan
Fixed Location
Freitag, Ruth
French Revolution
Fruin, Robert
Gessner, Conrad
Guiles, Kay
Gutenberg, Johannes
Heading
Hensen, Steve
Inventory
Jewett, Charles Coffin
Kilgour, Frederick
La Fontaine, Henri
Main Entry
Maunsell, Andrew
Mooers, Calvin
Muller, Samuel
Number Building
Original Order
Otlet, Paul
Panizzi, Anthony
Paris Principles
Parker, Ralph
Pinakes
Provenance
Ranganathan, S. R.
Reference
Respect des Fonds
Rostgaard, Frederik
Shaw, Ralph
Shelflist
Stereotyping
Subdivision
Taube, Mortimer
Trithemius, Johannes
Whytefield, John

Some Important Acronyms in This Chapter	
AACR:	*Anglo-American Cataloging Rules*
AACR2:	*Anglo-American Cataloguing Rules, 2nd Edition*
AAT:	*Art & Architecture Thesaurus*
ALA:	American Library Association
APPM:	*Archives, Personal Papers, and Manuscripts*
ASIS&T:	Association for Information Science and Technology
CCO:	*Cataloging Cultural Objects*
CRG:	Classification Research Group
DACS:	*Describing Archives: A Content Standard*
DDC:	*Dewey Decimal Classification*
EC:	*Expansive Classification*
IFLA:	International Federation of Library Associations and Institutions
ISBD:	*International Standard Bibliographic Description*
LC:	Library of Congress
LCC:	*Library of Congress Classification*
LCSH:	*Library of Congress Subject Headings*
MARC:	MAchine-Readable Cataloging
MARC-AMC:	MARC Format for Archival and Manuscripts Control
OCLC:	Online Computer Library Center
OSS:	Office of Strategic Service
PRECIS:	*Preserved-Context Indexing System*
RDA:	*Resource Description & Access*
UBC:	Universal Bibliographic Control
UDC:	*Universal Decimal Classification*

NOTES

All URLs accessed June 2017.

1. Lionel Casson, *Libraries in the Ancient World* (New Haven, CT: Yale University Press, 2001), 4.

2. Michael H. Harris, *History of Libraries in the Western World*, 4th ed. (Lanham, MD: Scarecrow, 1999), 19–20.

3. For more information about Callimachus's work, see Rudolf Blum, *Kallimachos: The Alexandrian Library and the Origins of Bibliography* (Madison: University of Wisconsin Press, 1991).

4. Martin Davies, "Medieval Libraries," *International Encyclopedia of Library Histories*, ed. David Stam (London: Routledge, 2015), 104.

5. Alain Besson, *Medieval Classification and Cataloguing: Classification Practices and Cataloging Methods in France from the 12th to 15th Centuries* (Biggleswade, UK: Clover Publications, 1980), 38.

6. Hugo Kunoff, *The Foundations of the German Academic Library* (Chicago: American Library Association, 1982), 10.

7. Montague Rhodes James, *The Ancient Libraries of Canterbury and Dover* (Cambridge: Cambridge University Press, 1903), lx.

8. Bodleian Library & Radcliffe Camera, "History of the Bodleian," University of Oxford, 2016, https://www.bodleian.ox.ac.uk/bodley/about-us/history.

9. Judith Hopkins, "The 1791 French Cataloging Code and the Origins of the Card Catalog," *Libraries and Culture* 27, no. 4 (Fall 1992): 379–83.

10. George Watson Cole, "An Early French 'General Catalog,'" *Library Journal* 25, no. 7 (July 1900): 330.

11. Mary Ellen Quinn, "Catalog Card," in *Historical Dictionary of Librarianship* (Lanham, MD: Rowman & Littlefield, 2014), 76.

12. The Austrian National Library claims that the first card catalog was actually created by Gottfried van Swieten in 1780. See Österreichische Nationalbibliothek "1780: The Oldest Card Catalogue," 2016, https://www.onb.ac.at/en/about-us/the-history-of-the-austrian-national-library/1780-the-oldest-card-catalogue/.

13. Joseph Smalley, "The French Cataloging Code of 1791: A Translation," *Library Quarterly* 61, no. 1 (January 1991): 5. Smalley provides a complete translation of the catalog code as distributed to the libraries, which were confiscated property of the crown in France and therefore under the decree of the National Constituent Assembly.

14. Ibid., 1–14.

15. Ruth French Strout, "The Development of the Catalog and Cataloging Codes," *Library Quarterly* 26, no. 4 (October 1956): 268.

16. Dorothy May Norris, *A History of Cataloguing and Cataloguing Methods, 1100–1850* (London: Grafton, 1939), 206.

17. Louis Fagan, *The Life of Sir Anthony Panizzi,* 2nd ed., reprint (New York: Burt Franklin, 1970), 134. (Originally published 1880.)

18. British Museum, Department of Printed Books, "Rules for the Compilation of the Catalogue," in *Catalogue of Printed Books in the British Museum*, ed. Anthony Panizzi, vol. 1 (London: British Museum, 1841), [v]–ix.

19. Harvard College Library, *Catalogus Librorum Bibliothecae Collegij Harvardini: quod est Cantabrigiae in Nova Anglia* (Boston: B. Green, 1723–1736), https://iiif.lib.harvard.edu/manifests/view/drs:42282481$5i.

20. Jim Ranz, *The Printed Book Catalogue in American Libraries, 1723–1900* (Chicago: American Library Association, 1964).

21. Personal communication between Arlene G. Taylor and Michael Carpenter. For further information about the drafts of Panizzi's rules, see Michael Carpenter, "The Original 73 Rules of the British Museum: A Preliminary Analysis," *Cataloging & Classification Quarterly* 35, no. 1/2 (2002): 23–36.

22. Charles Coffin Jewett, *On the Construction of Catalogues of Libraries, and Their Publication by Means of Separate, Stereotyped Titles,* 2nd ed. (Washington, DC: Smithsonian Institution, 1853), 18.

23. Elaine Svenonius, *The Intellectual Foundation of Information Organization* (Cambridge, MA: MIT Press, 2001), 79.

24. See the discussion of Cutter's "objects" of the catalog in Chapter 3.

25. Charles A. Cutter, *Rules for a Dictionary Catalog,* 4th ed. (Washington, DC: Government Printing Office, 1904; reprint, London: Library Association, 1962), 12.

26. Andrew Osborn, "The Crisis in Cataloging," in *Foundations of Cataloging: A Sourcebook,* eds. Michael Carpenter and Elaine Svenonius (Littleton, CO: Libraries Unlimited, 1985), 90–103. Also available in *Library Quarterly* 11, no. 4 (October 1941): 393–411.

27. Seymour Lubetzky, *Cataloging Rules and Principles: A Critique of the A.L.A. Rules for Entry and a Proposed Design for their Revision* (Washington, DC: Library of Congress, 1953).

28. Cutter, *Rules for a Dictionary Catalog*, 66–79.

29. American Library Association, *List of Subject Headings for Use in Dictionary Catalogs,* 2nd ed., rev. (Boston: Library Bureau, 1898), vi.

30. Library of Congress, "Classification Web," https://classificationweb.net/.

31. Library of Congress, “Library of Congress Subject Headings PDF Files,” https://www.loc.gov/aba/publications/FreeLCSH/freelcsh.html.

32. EBSCO, “Sears List of Subject Headings,” https://www.ebscohost.com/public/sears-list-of-subject-headings.

33. *Faceting* comprises small notations that stand for subparts of the whole topic (e.g., common aspects of topics, geography, form, language) and are strung together to create a complete classification notation. This is discussed further in Chapter 11.

34. Ernst Posner, “Some Aspects of Archival Development since the French Revolution,” *American Archivist* 3, no. 2 (April 1940): 159–72.

35. Samuel Muller, Johan A. Feith, and Robert Fruin, *Handleiding voor het ordenen en beschrijven van archieven* (Groningen: Erven B. van der Kamp, 1898).

36. Samuel Muller, Johan A. Feith, and Robert Fruin, *Manual for the Arrangement and Description of Archives*, translation of the second edition by Arthur H. Leavitt; with new introductions by Peter Horsman [and others] (Chicago: Society of American Archivists, 2003), 9.

37. Samuel Muller, Johan A. Feith, and Robert Fruin, *Manual for the Arrangement and Description of Archives*, drawn up by the Netherlands Association of Archivists (New York: H. W. Wilson, 1940).

38. Marjorie Rabe Barritt, “Coming to America: Dutch *Archivistiek* and American Archival Practice,” *Archival Issues* 18, no. 1 (1993). Reprinted in Muller, Feith, and Fruin, *Manual* (2003), xxxvi.

39. Ibid., l.

40. Steven L. Hensen, *Archives, Personal Papers, and Manuscripts: A Cataloging Manual for Archival Repositories, Historical Societies, and Manuscript Libraries* (Washington, DC: Cataloging Distribution Service, 1983).

41. Steven L. Hensen, *Archives, Personal Papers, and Manuscripts: A Cataloging Manual for Archival Repositories, Historical Societies, and Manuscript Libraries*, 2nd ed. (Chicago: Society of American Archivists, 1989).

42. Paul Bourcier, Heather Dunn, and the Nomenclature Task Force, *Nomenclature 4.0 for Museum Cataloging: Robert G. Chenhall's System for Classifying Cultural Objects* (Lanham, MD: Rowman & Littlefield/AASLH, 2015).

43. W. Boyd Rayward, “The Origins of Information Science and the International Institute of Bibliography/International Federation for Information and Documentation (FID),” *Journal of the American Society for Information Science* 48, no. 4 (1997): 289–300.

44. Vannevar Bush, “As We May Think,” *Atlantic Monthly* 176 (July 1945): 101–8. Also available: https://www.theatlantic.com/magazine/archive/1945/07/as-we-may-think/303881/.

45. For a quick and fun overview of Otlet's idea, originally discussed in his *Traité de Documentation: Le Livre Sur le Livre, Théorie et Pratique* (Brussels: Editiones Mundaneum, 1934), see the video clip, “Paul Otlet, Visioning a Web in 1934,” on YouTube, https://www.youtube.com/watch?v=hSyfZkVgasI. The clip is from the documentary by Françoise Levie and produced by Sofidoc Productions, *The Man Who Wanted to Classify the World* (New York: Filmakers Library, 2004).

46. Henriette Avram, “Defining the Fields of Data for the LC Machine-Readable Bibliographic Records,” as quoted in Karen M. Spicher, “The Development of the MARC Format,” *Cataloging & Classification Quarterly* 21, no. 3/4 (1996): 81.

47. Donna Rosenheck, “OCLC: From an Historical Perspective,” *The Katharine Sharp Review*, no. 4 (1997), https://www.ideals.illinois.edu/bitstream/handle/2142/78252/rosenheck_oclc.htm?sequence=3.

48. OCLC, “WorldCat,” http://www.oclc.org/en/worldcat.html.

SUGGESTED READINGS

Baker, Nicholson. "Discards." *New Yorker* 70, no. 7 (April 4, 1994): 64–86.

Berner, Richard C. "Historical Development of Archival Theory and Practices in the United States." *Midwestern Archivist* 7, no. 2 (1982): 103–17.

Burke, Frank G. "Archives: Organization and Description." In *World Encyclopedia of Library and Information Services,* 3rd ed., edited by Robert Wedgeworth, 63–68. Chicago: American Library Association, 1993.

Bush, Vannevar. "As We May Think." *Atlantic Monthly* 176 (July 1945): 101–8. Also available: https://www.theatlantic.com/magazine/archive/1945/07/as-we-may-think/303881/.

Carpenter, Michael. "The Original 73 Rules of the British Museum: A Preliminary Analysis." *Cataloging & Classification Quarterly* 35, no. 1/2 (2002): 23–36.

Denton, William. "FRBR and the History of Cataloging." In *Understanding FRBR: What It Is and How It Will Affect Our Retrieval Tools,* edited by Arlene G. Taylor, 35–57. Westport, CT: Libraries Unlimited, 2007.

Dunkin, Paul S. *Cataloging U.S.A.* Chicago: American Library Association, 1969. Chapter 1: "Mr. Cutter's Catalog" and Chapter 2: "The Prophet and the Law: Codes After Cutter."

Fishbein, Meyer H. "Archives: Records Management and Records Appraisal." In *World Encyclopedia of Library and Information Services.* 3rd ed., edited by Robert Wedgeworth, 60–63. Chicago: American Library Association, 1993.

Harris, Michael H. *History of Libraries in the Western World.* 4th ed. Lanham, MD: Scarecrow, 1995.

Hopkins, Judith. "The 1791 French Cataloging Code and the Origins of the Card Catalog." *Libraries & Culture* 27, no. 4 (Fall 1992): 378–404.

Humbert, de Romans. *Regulations for the Operation of a Medieval Library.* St. Paul: Associates of the James Ford Bell Library, University of Minnesota, 1980.

Jackson, Sidney L. *Libraries and Librarianship in the West: A Brief History.* New York: McGraw-Hill, 1974.

Joudrey, Daniel N., Arlene G. Taylor, and David P. Miller. *Introduction to Cataloging and Classification*. 11th ed. Santa Barbara, CA: Libraries Unlimited, 2015. Chapter 2: "Development of Catalogs and Cataloging Codes."

Lancaster, F. W. "Whither Libraries? or Wither Libraries." *College & Research Libraries* 39, no. 5 (September 1978): 345–57; reprinted in *College & Research Libraries* 50, no. 4 (July 1989): 406–19.

Osborn, Andrew D. "The Crisis in Cataloging." In *Foundations of Cataloging: A Sourcebook,* edited by Michael Carpenter and Elaine Svenonius, 90–103. Littleton, CO: Libraries Unlimited, 1985. Originally published in *Library Quarterly* 11, no. 4 (October 1941): 393–411.

Reynolds, Dennis. *Library Automation: Issues and Applications.* New York: Bowker, 1985. Chapter 4: "History: The Public Catalog."

Russell, Beth M. "Hidden Wisdom and Unseen Treasure: Revisiting Cataloging in Medieval Libraries." *Cataloging & Classification Quarterly* 26, no. 3 (1998): 21–30.

Smalley, Joseph. "The French Cataloging Code of 1791: A Translation." *Library Quarterly* 61, no. 1 (January 1991): 1–14.

Stoll, Clifford. *Silicon Snake Oil: Second Thoughts on the Information Highway.* New York: Doubleday, 1995, 197–203.

Strout, Ruth French. "The Development of the Catalog and Cataloging Codes." *Library Quarterly* 26 (October 1956): 254–75.

Taylor, Arlene G. "Cataloguing." In *World Encyclopedia of Library and Information Services.* 3rd ed., edited by Robert Wedgeworth, 177–81. Chicago: American Library Association, 1993.

Witty, Francis J. "The Beginnings of Indexing and Abstracting: Some Notes Towards a History of Indexing and Abstracting in Antiquity and the Middle Ages." *The Indexer* 8 (1973): 193–98.

* Much of the material in the first half of this chapter is based on Ruth French Strout, "The Development of the Catalog and Cataloging Codes," *Library Quarterly* 26, no. 4 (1956): 254–75.

CHAPTER 3

RETRIEVAL TOOLS

This chapter discusses retrieval tools, which are basic building blocks in the organization of recorded information. It addresses the following questions: Why do we need retrieval tools? What are the basic retrieval tools? What are their formats and their functions?

THE NEED FOR RETRIEVAL TOOLS

Retrieval tools are systems created for discovering information. They are designed to help users find, identify, select, obtain, and explore information resources of all types. They assist users by retrieving information directly (e.g., through a web search) or through relevant documents that are part of organized collections (e.g., a catalog search). At this writing, retrieval tools typically contain records that act as surrogates for resources. That is, each *surrogate record* (also called a *description* or *metadata*) provides enough information, such as creator, title, subject, and date, so that it can serve as a short representation for and facilitate access to an individual resource in a collection (e.g., a DVD in a library).

Surrogate records, however, may not be the only way to retrieve information. As Semantic Web technologies advance, the emphasis on the surrogate record as a unit of description may wane. Instead, more focus may be placed on creating individual metadata statements that are separately linked to a resource as a form of identification and description. At *this* time, however, the surrogate record remains the primary unit of description in most information environments and, thus, is still the focus of this chapter. Surrogate records are arranged and retrieved by *access points.* An access point can be a name, title, or subject term chosen by an indexer or cataloger. In online systems an access point can be almost any word in a record if keyword searching of every word that is not a stopword is allowed.

Retrieval tools are essential as basic building blocks for a system that will organize as much of the world's recorded information as possible. A dream of being able to provide access to all recorded information has existed since 1892, when Paul Otlet and Henri La Fontaine organized a conference to create *Universal Bibliographic Control* (UBC). They wanted to create a central file that would include surrogate records particularly for scientific articles in all the scientific journals of the world. The magnitude of the undertaking meant that new techniques different from conventional library practices had to be developed. As UBC evolved throughout the twentieth century, it came to mean that each country of the world would be responsible for creating descriptions for its information resources and would share those records with all other countries. The concept was extended also to authority control of names and titles used as access points. A program of the International Federation of Library Associations and Institutions (IFLA) combined the ideals of UBC with the concept of making the records machine readable. The many retrieval tools that have been developed as a result of Otlet and La Fontaine's dream have brought us closer to UBC.

THE BASIC RETRIEVAL TOOLS, THEIR FORMATS, AND THEIR FUNCTIONS

The basic retrieval tools discussed in this chapter are

- bibliographies

- catalogs
- indexes
- finding aids
- registers
- search engines

Databases and bibliographic networks have distinct roles in housing retrieval tools. They are discussed in the next chapter, which addresses the topics of systems and system design.

Bibliographies

Fundamentally, bibliographies are lists of resources. They are essential to scholars and to those involved professionally with books and other sources of information (e.g., collectors, dealers, librarians), and they are also useful resources for all serious readers. They bring together lists of sources based on subject matter, creator, time period, and the like (see the more detailed list of these below). Some bibliographies include *annotations,* that is, brief reviews indicating the subject matter and/or commenting on the usefulness of the information resource.

Bibliographies can be attached to a scholarly work and consist of the information resources that were consulted by the author of the work, or they can be completely separate entities—works in their own right. Each resource represented in the list has a short description referred to as a *citation* (not to be confused with an *annotation*). A typical citation includes creator, title, edition or version, publisher, place, and date of publication for a book or other such whole entity. For a part of a work (such as a journal article or a poem), one typically includes creator, title, name of the larger work, volume and issue numbers (if applicable), date, and page numbers or other part designation. Some citations also include physical characteristics.

In a bibliography each citation usually appears in only one place, most often under the creator (first creator if more than one) of the work. This is an example of a retrieval tool that generally provides only one access point. In a bibliography that is arranged only by creators' names, for example, other attributes such as titles, second or later creators, and subjects are not access points (i.e., users who only know the title, the name of a second or later creator, or the subject of a work will not easily be able to retrieve information about it from the bibliography). The citations may be constructed according to various styles, one of which is chosen by the creator of the bibliography or by a publisher (for bibliographies attached to scholarly works). Examples of style manuals and a citation from each are provided below:

- **APA (American Psychological Association)**[1]
  Mitchell, T. R., & Larson, J. R., Jr. (1987). *People in organizations: An introduction to organizational behavior* (3rd ed.). New York: McGraw-Hill.

- **Chicago Manual of Style**[2]
  Mitchell, Terence R., and James R. Larson, Jr. *People in Organizations: An Introduction to Organizational Behavior.* 3rd ed. New York: McGraw-Hill, 1987.

- **MLA (Modern Language Association)**[3]
  Mitchell, Terence R., and James R. Larson, Jr. *People in Organizations: An Introduction to Organizational Behavior.* 3rd ed. New York: McGraw-Hill, 1987. Print.

- **Science (Scientific Style and Format)**[4]
  Mitchell TR, Larson JR, Jr. People in organizations: an introduction to organizational behavior. 3rd ed. New York (NY): McGraw-Hill; 1987.

- **Turabian**[5]
  Mitchell, Terence R., and James R. Larson, Jr. *People in Organizations: An Introduction to Organizational Behavior.* 3rd ed. New York: McGraw-Hill, 1987.

- **Style Manual (U.S. Government Style Manual)**[6]
  Mitchell, Terence R., and James R. Larson, Jr. *People in Organizations: An Introduction to Organizational Behavior.* 3rd ed. (New York: McGraw-Hill, 1987).

The above examples are for a book, but each style manual provides guidelines for citing all types of resources, including journal articles, theses and dissertations, and websites, to name a few.

Each bibliography has a particular focus or arrangement. The most common include the following:

Focus	Purpose	Example
Subject	**Gathers together resources about a particular subject.**	***A Bibliography of Female Economic Thought Up to 1940***[7]
Creator	**Gathers all or some works of a particular creator and sometimes including sources about the creator.**	***A Bibliography of Jane Austen***[8]
Language	**Gathers textual resources in which the text is in one or more particular languages.**	***An Extensive Bibliography of Studies in English, German, and French on Turkish Foreign Policy, 1923–1997***[9]
Time Period	**Gathers all resources that came to light in a particular time period.**	***Russian and Soviet Education 1731–1989***[10]
Locale	**Gathers resources created in a particular location. This could be a large locale such as a whole continent, or a smaller locale such as a country, region, state, city, or community. It could also be an institution, such as a bibliography of all the works of the faculty in a particular university.**	***Area Bibliography of China***[11] **University of Iowa College of Law's *Faculty Bibliography***[12]
Publisher	**Gathers all of the products of a particular publisher**	***The Stinehour Press: Work of the First Fifty Years***[13]
Form	**Gathers information resources that appear in a certain form, format, or genre (e.g., videocassettes, electronic resources, poetry, biographies). These are virtually always combined with one of the other foci.**	***Maps and Mapping of Africa: A Resource Guide***[14]

Two or more of these foci are often combined in bibliographies. For example, the title above that illustrates *language* is a combination of language, subject, and time period; the one illustrating *time period* is a combination of subject, locale, and time period; and the one illustrating *form* is a combination of form and locale.

There is a special kind of bibliography found in libraries that is truly meant to be a retrieval tool. They were for many years called *pathfinders*, but now they are usually referred to as *library guides*, *subject guides*, or *research guides*. These tools are subject bibliographies with a special function in the library world. While most bibliographies do not indicate whether a location actually has a specific resource, research guides focus on the resources in a defined subject area available in a particular setting. In addition to the list of relevant resources, the guide may also include a list of locally accessible databases to search, specific instructions on how to search the local catalog, and a list of specific reference sources related to the particular subject area. For example, the guide may provide a list of relevant subject headings that may be used in the local catalog. Subject and reference librarians have created these tools for decades. Originally pathfinders were printed documents available only in the library, but now these guides are primarily found on a library's website.[15] In recent years, software products have been developed for the creation and management of library research guides. These range from proprietary software products (such as LibGuides)[16] to open-source software (such as SubjectsPlus).[17] Another option is to adapt other web tools, such as wikis or blogs, to meet the purpose, as they are somewhat inexpensive and easy to implement.

Catalogs

Catalogs provide access to individual items within collections of information resources (e.g., physical entities such as books, DVDs, and maps in a library; artists' works in an art museum; web pages). Each resource is represented by a description that is somewhat longer than an entry in a bibliography. The descriptions are assigned one or more access points. As mentioned earlier, an access point can be almost any word in a record when keyword searching is used; however, the term *access point* is usually applied to a particular name, title, or subject that is listed on the record separately from the description. An access point is constructed in a certain order (e.g., surname followed by forename or forenames), and it is maintained under authority control. *Authority control* is the process of bringing together all of the forms of name that apply to a single entity; gathering together all the variant titles that apply to a single work; and relating all the synonyms, related terms, broader terms, and narrower terms that apply to a particular subject heading.

The descriptions in a catalog are constructed according to a standard style selected by a particular community (e.g., *RDA: Resource Description & Access* for libraries; *Describing Archives: A Content Standard*

(DACS) for archives in the United States; *Cataloging Cultural Objects* (CCO) for some art collections). Several different standards for description are discussed in more detail in Chapter 7.

Purposes of Catalogs

Catalogs have traditionally served two main user groups. One group is the employees of the institution, who need to retrieve information resources or who need to retrieve information about those resources (i.e., metadata) in order to work with the collection. For example, the catalog is used by collection development librarians in their process of discovering what the library already owns or does not own before selecting new materials; another such use is by the employee of a museum who is looking for objects to place in an exhibit.

The most commonly understood use for a catalog, though, is use by the patrons of the institution who wish to borrow material or make use of it on the premises. If such users have a known resource in mind, they may search for it in the catalog by creator or title (called *known-item searching*). Users might also try searching by keyword if they only remember certain words of the title or some other attribute. If users know they want resources by a particular creator, they may search under the creator's name. If users do not know of a particular resource but are searching for something on a particular topic, they may use a subject heading search, a subject keyword search, or a general keyword search. In online catalogs, keyword searches are often appropriate for helping a person find a record that looks like it might be on the user's topic. Once a potentially useful record is found, the user can identify in that record an authority-controlled subject heading or a classification notation for the topic of interest. The user, then, may conduct a search for the subject heading or the classification notation in the catalog (often by clicking on hyperlinks in the catalog). Alternatively, one may go to the location of the classification notation in the stacks to determine if there are other pertinent works shelved with the one that has been identified.

A number of attempts have been made through the years to identify purposes of catalogs. Charles A. Cutter presented his "objects" (i.e., objectives) of a catalog in his *Rules for a Printed Dictionary Catalog* in 1876. Cutter's objectives are listed in Figure 3.1.[18] If the language of the objectives were updated, they would now, among other things, refer to *resources*, not just to *books*. The first object, a finding function, says that a catalog should help a user retrieve a resource if one of its key access points is known and it is part of the collection. The second object, a gathering function, states that a catalog should be able to display *all* resources in the collection by a creator, on a specific subject, or in a particular genre or literary form. The third object, a selecting function, is to help users to choose the resource they need based on bibliographic attributes (e.g., version, language, format, date) or the nature of the content (e.g., genre, subject matter). Cutter was speaking only of library catalogs in which books were represented, but if these objectives are broadened to archives, museums, and other collecting institutions, they still represent what catalogs and other retrieval tools are supposed to do today.

Objects

1. To enable a person to find a book of which either
   - (A) the author
   - (B) the title } is known.
   - (C) the subject

2. To show what the library has
   - (D) by a given author
   - (E) on a given subject
   - (F) in a given kind of literature.

3. To assist in the choice of a book
   - (G) as to its edition (bibliographically).
   - (H) as to its character (literary or topical).

Figure 3.1. Cutter's Objects of the Catalog.

Seymour Lubetzky worked in the mid-twentieth century to simplify cataloging rules, and he posited that the cataloging code should be reconstructed "in accordance with deliberately adopted objectives . . . and well considered principles."[19] He then stated that:

> The first objective is to enable the user of the catalog to determine readily whether or not the library has the book he wants. . . . The second objective is to reveal to the user of the catalog, under one form of the author's name, what works the library has by a given author and what editions or translations of a given work.[20]

Lubetzky's work was the basis for the Paris Principles, adopted at the International Conference on Cataloguing Principles in Paris in 1961. The second of the principles was for "Functions of the Catalogue" and stated that the catalog should be an efficient instrument for ascertaining

2.1. whether the library contains a particular book specified by
- (a) its author and title, *or*
- (b) if the author is not named in the book, its title alone, *or*
- (c) if author and title are inappropriate or insufficient for identification, a suitable substitute for the title; and

2.2. (a) which works by a particular author *and*
- (b) which editions of a particular work are in the library.[21]

In 1998 IFLA published *Functional Requirements for Bibliographic Records* (FRBR), which identified four generic user tasks that the bibliographic record is intended to support. These four tasks have been built on the work of Cutter, Lubetzky, and the Paris Principles, and essentially represent the four main functions of a catalog:

- to *find* entities that correspond to a user's search criteria
- to *identify* entities (e.g., persons, works, subjects)
- to *select* entities appropriate to the user's needs
- to *obtain* access to the entities described[22]

The fourth user task in FRBR actually is an addition to the functions identified by Cutter, but it was recognized by several writers in the late twentieth century that a catalog differs from certain other retrieval tools in that it facilitates physically locating the information resources that are represented by descriptions in the catalog.[23] In 2016 a fifth user task was proposed in the draft IFLA Library Reference Model (LRM)—a document that attempts to harmonize FRBR and two other functional-requirements models established by IFLA. The fifth task is to *explore* relationships between one resource and another.[24] The LRM has not been adopted at the time of this writing, but it is expected that this addition to the user tasks will be approved.

Another important purpose served by catalogs has traditionally been to act as an inventory of the collection—that is, to provide a record of what is owned. Often a *shelflist* has been used to accomplish this purpose. A shelflist includes one copy of every record in a catalog arranged in the order in which the information resources, objects, and so forth, are arranged on the shelf. So, for example, one could use the shelflist in the process known as *shelf reading* to determine whether resources from the collection were misshelved. Originally, shelflists were literally in the order of items on the shelf, starting a new sequence with each change in format, or change in collection location, or change in size. Later, shelflists often were arranged by classification notation regardless of format or other categorization. This is the way the concept works in online catalogs. The purpose of serving as an inventory is still there, but the mental image of a shelflist with arrangement as it is on the shelf is lost.

A *union catalog* is a variation of the concept that a catalog represents just the holdings of *one* institution. A union catalog represents the holdings of *more than one* institution or collection. For example, a union catalog can show the resources held in a main library and also in one or more of its branch libraries. The location information indicates where a resource is held. A union catalog of a consortium of institutions works the same way: location information shows which resources are housed in which of the cooperating institutions. The ultimate union catalog is the one maintained by the largest bibliographic network, OCLC (see description in Chapter 4), where each information resource has one master record; associated with that record is a holdings record that shows the holding symbol of each member of the network that has cataloged the resource through OCLC or has asked that its symbol be added to the record.

Forms of Catalogs

Catalogs can appear in different forms and can have different arrangements. The following formats are discussed:

- book catalogs
- card catalogs
- microform catalogs
- online catalogs

Book Catalogs. Book catalogs originally were just handwritten lists. After the invention of printing with moveable type, book catalogs were printed, but not always in a discernible order. Eventually, entries were printed in alphabetical or classified order, but book catalogs were very expensive to produce and were, therefore, not reproduced and updated often. By the early 1900s, book catalogs were almost completely replaced with card catalogs that could be updated as soon as the cards could be filed and were relatively inexpensive to maintain and create. Book catalogs had a brief renaissance in the 1960s and 1970s when (1) computers again made them easy and less expensive to produce, (2) large card catalogs became unwieldy, and (3) rapid growth of new libraries and new branches made it desirable to have multiple copies of catalogs. But in order to keep a book catalog up to date, supplements were usually produced, resulting in multiple look-ups for one search. In addition, it was usually three to six months before supplements were produced, meaning that new materials were not represented in the catalog during that time.

Online catalogs have replaced both book and card catalogs in most technologically advanced and economically advantaged countries, but book catalogs are still used in some cases for rare materials, catalogs of exhibits, artists' works, and so forth. Book catalogs provide a way to make the contents of special collections known to users in many locations. For example, book catalogs of historical societies are popular acquisitions by collectors of genealogical materials. This use, though, may be replaced by availability of such catalogs online. Book catalogs also still exist in some libraries and archives as the only retrieval tool available to access older materials, because some institutions have chosen not to convert all of the catalog records for rarely used resources into machine-readable form for the online catalog.

An advantage of book catalogs over online catalogs once was that book catalogs are compact and portable and can be consulted anywhere they can be carried. However, with laptop computers, smart phones, and wireless technology, online catalogs are now accessible from almost anywhere. Another advantage of book catalogs is that glancing over a page of book catalog entries is relatively fast, and some people prefer this to paging through screen after screen of online responses. An illustration of a book catalog is found in Figure 3.2.

Cambridgeport Free Library
Dictionary Catalog

HV9644 .H8 1887	**Jottings from jail** : notes and papers on prison matters / by the Rev. J. W. Horsley. – London : T. F. Unwin, 1887. – viii, 259 pages ; 19 cm
PQ2311 .J73 A22 1983	**Joubert, Joseph, 1754-1824.** The notebooks of Joseph Joubert : a selection / edited and translated by Paul Auster ; with an afterword by Maurice Blanchot. – San Francisco : North Point Press, 1983. – x, 181 pages ; 23 cm. – Translation of: Les carnets de Joseph Joubert.
Z693 .W94 2015	**Joudrey, Daniel N.** Introduction to cataloging and classification / Daniel N. Joudrey, Arlene G. Taylor, and David P. Miller. – Eleventh edition. – Santa Barbara, California : Libraries Unlimited, 2015. – xxv, 1048 pages : illustrations ; 26 cm. – Library and information science text series. – Includes bibliographical references (pages 1001-1022) and index.
Z666.5 .T39 2009	The organization of information / Arlene G. Taylor and Daniel N. Joudrey. – Third edition. –Westport, Conn. : Libraries Unlimited, 2009. – xxvi, 512 pages : illustrations ; 26 cm. – Library and information science text series. – Includes bibliographical references (pages 479-498) and index.
PQ2635 .A25 J6 1925	**Les joues en feu** : poèmes anciens et poèmes inédits 1917-1921 / Raymond Radiguet. – Paris : B. Grasset, 1925. – 104 pages : 1 portrait ; 19 cm
E237 .J6 H34 1973	**Jouett, Jack, 1754-1822—Juvenile fiction.** Jack Jouett's ride / written and illustrated by Gail E. Haley. – 1st edition. – New York : Viking Press, ©1973. – 31 unnumbered pages : color illustrations ; 23 x 29 cm
PQ2635 .O9628 A6 1966	**Jouffroy, Alain, 1928-** Saint-Pol-Roux / édité par Alain Jouffroy. – Paris : Mercure de France, 1966. – xxxv, 296 pages ; 21 cm. – Includes bibliographical references (pages 285-290).
LB2332 .J68 1967	**Joughin, Louis.** Academic freedom and tenure : a handbook of the American Association of University Professors / edited by Louis Joughin. – Madison : University of Wisconsin Press, 1967. – xiv, 343 pages : illustrations ; 22 cm
KF224 .S2 J6 1978	The legacy of Sacco and Vanzetti / by Louis Joughin and Edmund M. Morgan ; introduction by Arthur M. Schlesinger. – Princeton, N.J. : Princeton University Press, 1978, ©1948. – xii, 596 pages ; 23 cm
PN1997 .J68 C35 1970	**Le jour se lève** / a film by Marcel Carné and Jacques Prévert. English translation and description of action by Dinah Brooke and Nicola Hayden. – New York : Simon and Schuster, 1970. – 128 pages : illustrations ; 21 cm
	Jouve, Nicole Ward ***See*** Ward Jouve, Nicole

237

Figure 3.2. A Representation of a Page from a Book Catalog.

Card Catalogs. In a card catalog, each bibliographic description is prepared on a standard 7.5 × 12.5 cm card (roughly 3 × 5 in.), and interfiled into drawers full of cards in specially designed cabinets. The dimensions of the cards were not always standard; in the early days of card catalogs, different sizes were in use (3 × 5 in., 1

1⁄2 × 5 in., 1 1⁄2 × 10 in., and 4 × 6 in.). Card catalogs were popularized in the United States by Library of Congress (LC) cards, first made available for sale in 1901, and by H. W. Wilson cards, which began production in 1938 in response to the needs of small libraries. (Both have now ceased card production.)

Technological advances encouraged further use of the card catalog. The descriptions on cards were at first hand-written, using a standardized *library hand* type of writing developed by Melvil Dewey in the 1870s (See Figure 3.3). Later, typewriters, one of the great technological innovations in libraries at the turn of the twentieth century, made handwritten cards unnecessary. Offset printing was used by LC for its cards. Then photocopying allowed the local creation of whole card sets from one master card. Finally, the advent of computer printing made it possible to have customized cards made either locally or at a distant facility. When created at a distance, the cards often were shipped to the receiver in boxes already alphabetized and ready to file.

Because of the influence of LC cards, card catalogs and the order of information on cards in libraries were standardized for many decades. Users of card catalogs could go from library to library, using catalogs with confidence that they would be able to use distant catalogs with as much ease as their local ones, although the filing arrangement of those catalogs often differed from library to library.

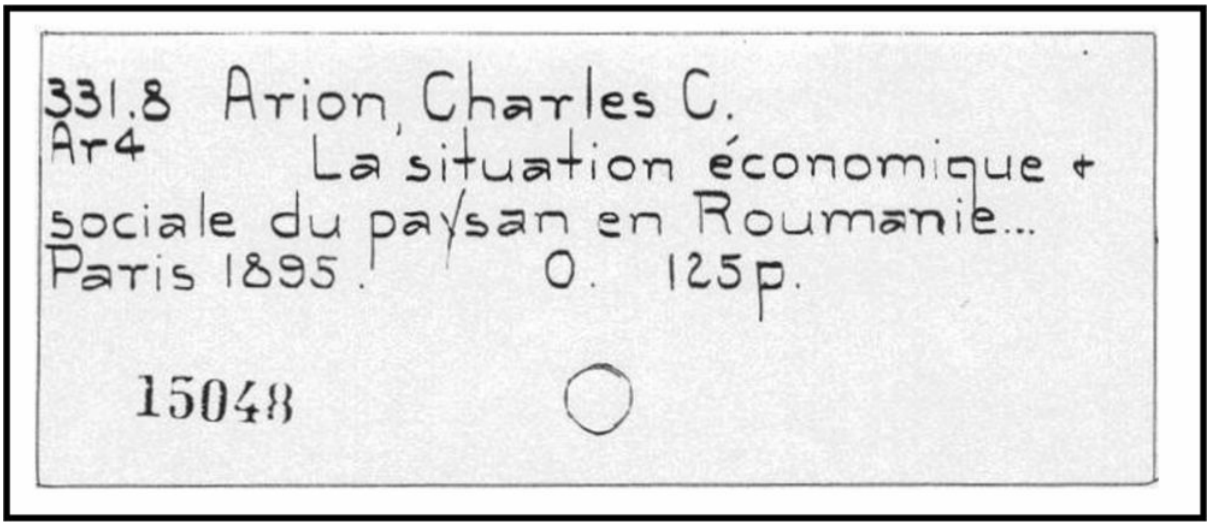

331.8 Arion, Charles C.
Ar4 La situation économique + sociale du paysan en Roumanie...
Paris 1895. O. 125p.
15048

Figure 3.3. Example of an 1895 Catalog Card Written in Library Hand.

Online catalogs have replaced most card catalogs in the United States, but some libraries, archives, and museums still have card catalogs, especially small institutions or those where there has been only minimal conversion of data to machine-readable form. In many other countries there are more card catalogs, and in some, card catalogs are still the predominant form. Figure 3.4 is an example of a more contemporary catalog card.

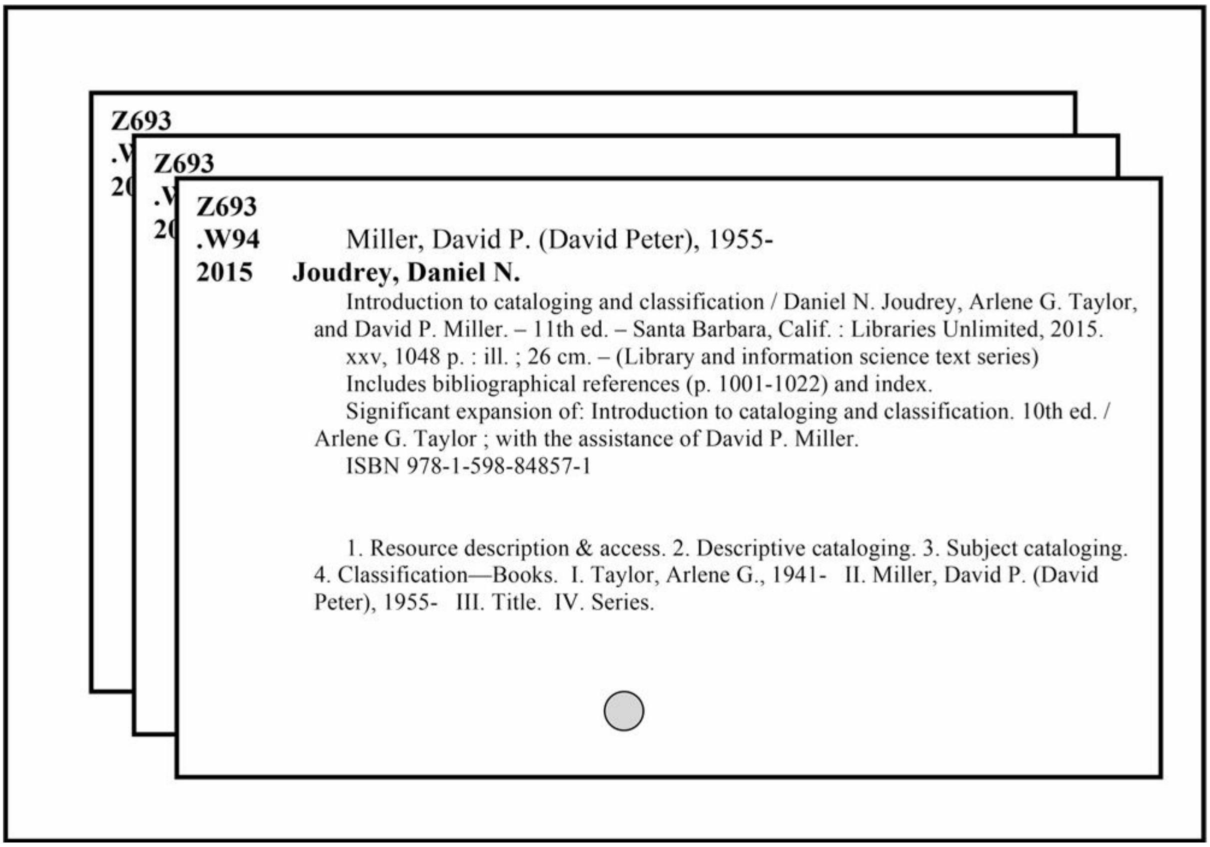


Figure 3.4. A Representation of a Printed Catalog Card.

Microform Catalogs. The creation of *computer output microform* catalogs (or COM catalogs) became possible in the 1960s. They are produced on either microfiche or microfilm and require a microform reader in order to be able to use them. They are produced like book catalogs, but because they do not have to be reproduced on paper and bound, they can be completely reproduced with new additions every three months or so without having to go through the supplement stage. Due to unpopularity, microform catalogs were replaced rather quickly by online catalogs. It has been found that users will use microfilm if that is the only way to get the information they need, but most people find the readers hard to use and difficult to read.

Online Catalogs. Online catalogs, often referred to as *online public access catalogs* or OPACs, are the predominant form of catalog in the United States and in a number of other countries today. In these catalogs, records are stored on a local or remote server. The records are displayed only as needed. There is much flexibility in the look of the displays. Figure 3.5 shows the basic search screen of the OPAC used at Simmons College's Beatley Library. Figure 3.6 shows the view of a single record in that library's catalog.

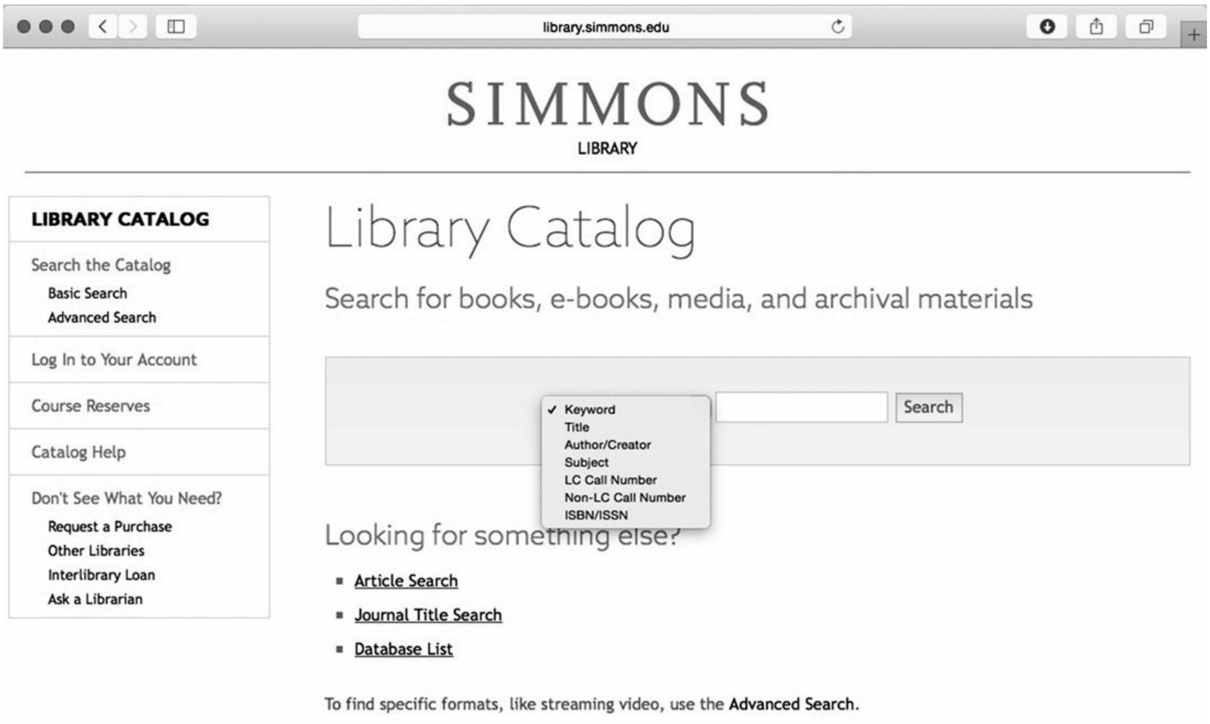
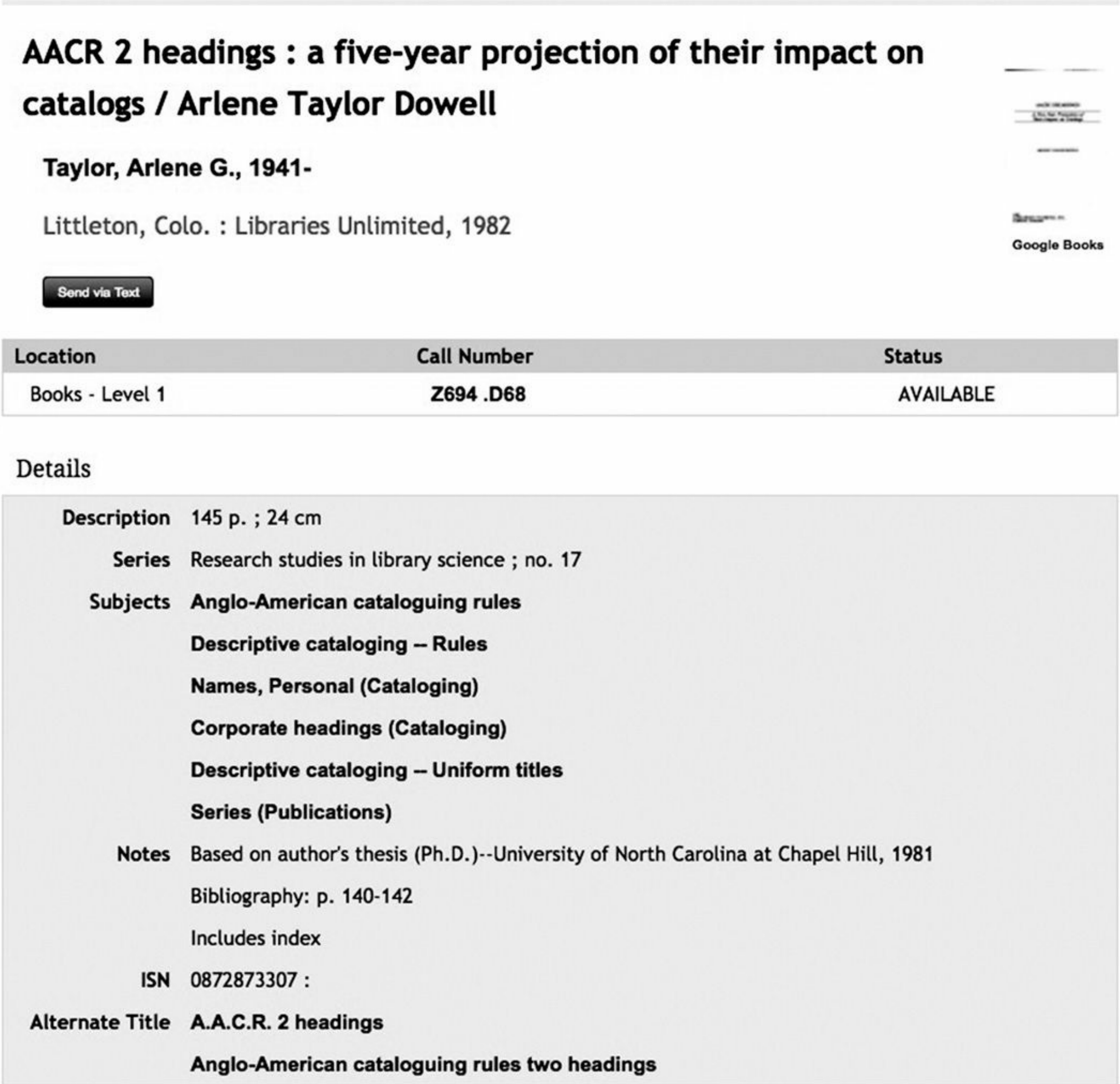


Figure 3.5. Basic Search Screen in an OPAC. (Source: Beatley Library, Simmons College, Boston, Massachusetts)

Online catalogs have not been standardized, although in two or more institutions that have purchased the same system, the displays may look somewhat similar. Writers in the field have long called for standardization, so that patrons can move from catalog to catalog or search multiple catalogs from the same location and find records displayed in the same manner, but this has yet to happen.[25] In recent years, dissatisfaction with online catalogs has been expressed by many in the library and information professions. This frustration—that the online catalog is still difficult to use—is intensified by comparisons to Internet search engines such as Google and Bing that are relatively easy to use. At times, this discontent has led some to question the long-term prospects for the online catalog. If catalog vendors respond by improving arrangements of displayed responses (e.g., by emphasizing relationships among resources), online catalogs may again be seen as providing something not available in search engines and then may survive. Some believe that when the Semantic Web is realized and all types of resources (bibliographic and otherwise) are connected using linked data techniques, then perhaps general search engines will replace the OPACs that we use today to find information in libraries. This, however, may be a long time coming.



AACR 2 headings : a five-year projection of their impact on catalogs / Arlene Taylor Dowell

Taylor, Arlene G., 1941-

Littleton, Colo. : Libraries Unlimited, 1982

Send via Text

Google Books

Location	Call Number	Status
Books - Level 1	**Z694 .D68**	AVAILABLE

Details

Description	145 p. ; 24 cm
Series	Research studies in library science ; no. 17
Subjects	**Anglo-American cataloguing rules**
	Descriptive cataloging -- Rules
	Names, Personal (Cataloging)
	Corporate headings (Cataloging)
	Descriptive cataloging -- Uniform titles
	Series (Publications)
Notes	Based on author's thesis (Ph.D.)--University of North Carolina at Chapel Hill, 1981
	Bibliography: p. 140-142
	Includes index
ISN	0872873307 :
Alternate Title	**A.A.C.R. 2 headings**
	Anglo-American cataloguing rules two headings

Figure 3.6. A Bibliographic Record in an OPAC. (Source: Beatley Library, Simmons College, Boston, Massachusetts, http://library.simmons.edu/record=b1055231~S0)

Arrangement and Displays within Catalogs

The records within catalogs must be arranged in some fashion or they are unusable. In card catalogs, records are arranged by being filed in a certain order. In book and microform catalogs, records are arranged by being printed one after another in a particular order. Records in OPACs, however, are not set in one place; there is flexibility because the records are arranged internally within the database either in sequence of order of entry into the system or in random order. This allows for the customization of the display of search results in online catalogs. So the arrangement discussed here applies to the arrangement of the *displayed* responses to search queries in the case of online catalogs and not to the internal order within the database. (For more information about arrangement of metadata displays, see Appendix C.)

In general, there were two basic arrangements that made sense in *printed catalogs* (i.e., all the formats of catalogs occurring before the online version). Catalog records were either in classified order or in alphabetical order. These arrangements are possible in the online catalog, although they are no longer the *only* choices for display. In today's OPACs, many options for displaying search results are offered; users may choose the arrangement that works best for their particular searches from a list of options. In the next sections, the two traditional approaches to the arrangement of printed catalogs are discussed; a brief section then follows on the choices for display in the online catalog.

Classified. Classified catalogs usually have more than one section. In what is considered to be the "main" section of the catalog, the arrangement or display is in the order of the classification scheme used in that institution. That is, this section is arranged in subject order, where the subject is represented by a classification notation rather than by subject terminology. For example, a book of recipes for use in a slow cooker is represented and arranged by the Dewey Decimal Classification (DDC) notation **641.5884** rather than by its analogous subject heading, **Electric cooking, Slow**. There can be as many classification notations assigned to a single record as there are subject concepts in the information resource.

Classified catalogs have the advantage that users can look at records for broader and narrower concepts at the same place they are looking for records on a specific concept. In a way, it is similar to browsing in the stacks with a classification notation, except that in the case of the classified catalog, each information resource is represented by several notations signifying all of its concepts, not just one notation as is true of resources in the stacks. An example of a page from a classified printed catalog is found in Figure 3.7; in it one can see the entries arranged by the call number associated with each resource.

As it is nearly impossible for anyone to know all the classification notations relating to a subject, there should be a section of the catalog that lists verbal representations of all topics and gives the notation for each topic (e.g., **Electric cooking, Slow** = **641.5884**). In some situations, this need could be met by placing a copy of the classification scheme at the catalog, because most classification schemes include an index of topics in alphabetical order. And, of course, there are users who want to search for authors and/or titles, so there are other sections of the catalog for these. In printed catalogs, these subject-term, author, and title sections are arranged in alphabetical order. In OPACs they are word-searchable and the order of displays varies.

Cambridgeport Free Library
Classified Catalog

641.3
R523w

Richardson, Joan.
Wild edible plants of New England : a field guide : including poisonous plants often encountered / by Joan Richardson. – Yarmouth, Maine : DeLorme Publishing, 1981. – x, 217 pages, 32 pages of plates : illustrations (some color) ; 23 cm
Includes bibliographical references (pages 189-190) and index.
ISBN 0-993301-09-6

641.303
T373n

Thayer, Samuel.
Nature's garden : a guide to identifying, harvesting, and preparing edible wild plants / by Samuel Thayer – Birchwood, Wisconsin : Forager's Harvest Press, 2010. – 512 pages : color illustrations ; 23 cm
Subtitle from cover.
Includes bibliographical references (pages 481-492) and index.
ISBN 978-0-9766-2661-9

641.58
D417b

Denzer, Kiko.
Build your own earth oven : a low-cost, wood-fired mud oven, simple sourdough bread, perfect loaves / Kiko Denzer, with Hannah Field. – 3rd edition. – Blodgett, Oregon : Hand Print Press, 2007. – 129 pages : illustrations (some color) ; 26 cm
"Revised, expanded, updated" – Cover.
Includes bibliographical references (pages 122-124) and index.
ISBN 978-0-96798-467-4

641.58
G618o

Goldenson, Suzanne, 1944-
The open-hearth cookbook : recapturing the flavor of early America / by Suzanne Goldenson with Doris Simpson. – Revised edition. – Chambersburg, PA : Alan C. Hood & Co., 2006. – ix, 164 pages : illustrations ; 23 cm
ISBN 978-0-911469-26-4

641.58
Y28g

Yarnell, Elizabeth.
Glorious one-pot meals : a revolutionary new quick and healthy approach to Dutch-oven cooking / Elizabeth Yarnell. – New York : Broadway Books, 2009. – xii, 223 pages ; 21 cm
Previously published: Denver, Colorado : Pomegranate Consulting, 2005.
Includes index.
ISBN 978-0-76793-010-9

641.5884
P558s

Phillips, Diane.
Slow cooker : the best cookbook ever with more than 400 easy-to-make recipes / Diane Phillips ; photographs by James Baigrie. – San Francisco : Chronicle Books, 2009. – 543 pages : illustrations (some color) ; 24 cm
"Glorious meals at the flip of a switch. The slow cooker is your new best friend for delicious breakfasts, lunches, dinners, and desserts" -- Page 4 of cover.
ISBN 978-0-81186-657-6

498

Figure 3.7. A Representation of a Page from a Classified Book Catalog.

Classified catalogs have traditionally been used in European and other countries where several languages are spoken and represented in the setting. The classified subject section of the catalog can include records for every language represented, with several languages being interfiled at the same classification notation. The sections of the catalog that give access to subject terms, authors, and titles can be in the language(s) appropriate to the clientele. The United States has not traditionally felt the need for this approach. Among the reasons the concept is still relevant today is that there is now global access to online catalogs, and a classified catalog can hold and display records in any language with classification notations that are universal. If indexes to a classified catalog are made in many languages, access can be gained through any one of the many languages. In addition, it makes broadening and narrowing of searches, as well as browsing, much easier.

Alphabetical. Early American catalogs were arranged by broad subject categories in alphabetical order. With a collection consisting of a few books, there was little need for elaborate classification or arrangement. As catalogs grew, the broad categories needed to be subdivided, so somewhat narrower categories were created and placed alphabetically within each broad category. For example, if the broad category were **Domestic animals**, the sub-categories under it might be **Cats**, **Cows**, **Dogs**, **Horses**, **Mules**, and so forth. These were called *alphabetico-classed catalogs.* As subject categories multiplied, it became more difficult to predict the subject category and where it would be found. It began to make sense to place *all* categories and sub-categories in alphabetical order. Cutter was instrumental in the development of what he called the *dictionary catalog.* He recommended alphabetical arrangement with authors, titles, and subjects all interfiled in the same catalog. An example of a printed dictionary catalog page is found earlier in this chapter in Figure 3.2. The page in that figure contains the following mix of entries:

1. a title entry for *Jottings from Jail*
2. an author entry for Joseph Joubert's notebooks
3. an author entry for Daniel N. Joudrey for a book on cataloging
4. another author entry for Joudrey for an earlier edition of this book, *The Organization of Information*
5. a title entry for a book of French poems, *Les Joues en Feu*
6. a subject entry for a biographical children's book about Jack Jouett
7. an entry for Alain Jouffroy, an editor
8. an entry for a handbook edited by Louis Joughin
9. an author entry for Joughin's book on Sacco and Vanzetti
10. a title entry for a book about the film *Le Jour se Lève*
11. a personal name cross-reference (from Jouve to Ward Jouve)

The interfiling of the various types of entries is what defines this type of arrangement as a *dictionary* catalog.

The dictionary card catalog was the standard for the first half of the twentieth century. Later, it became quite complicated to file new cards into large catalogs because of the complexity of filing rules that seemed to grow as the catalogs grew. Early filing rules had reflected the influence of the classified catalog and presorted catalog entries into categories. Cutter recommended alphabetical filing, but names and titles beginning with a word that could also be a topical subject (e.g., Glass) were supposed to be filed in the order: personal name, geographic entity, corporate body, subject heading, title. This is seen in the following example:

Glass, Richard	[personal name]
Glass Mountains (Tex.)	[geographic entity]
Glass Art Society	[corporate body]
Glass	[subject heading]
Glass art	[subject heading]
Glass menagerie	[title]

It can be seen that these are *not* alphabetical and it is easy to see how complicated such an arrangement was in large catalogs.

Attempts to break up the large files resulted in *divided catalogs.* Divided catalogs were alphabetical catalogs that were partitioned into two or possibly three sections. If there were two, they were often divided so that authors and titles were in one part and subject headings were in the other. In a catalog with three sections, the division was usually authors, titles, and subjects. In this tripartite arrangement, records for resources *about* a person were usually filed in the subject section, while records for works *created by* that person were filed in the author section. However, it was considered useful to keep records for works by and about a person together, so sometimes all *names* were placed in one part of the catalog whether it was for the name *as a creator* or for the name *as a subject*. In a divided card catalog, there might be separate cabinets (or sets of cabinets) used for the different sections of the catalog (see Figure 3.8). The divided catalog was easier and less expensive for the keepers of the catalog, but it was assumed that users knew the differences among author, title, and subject entries, which was not always the case.

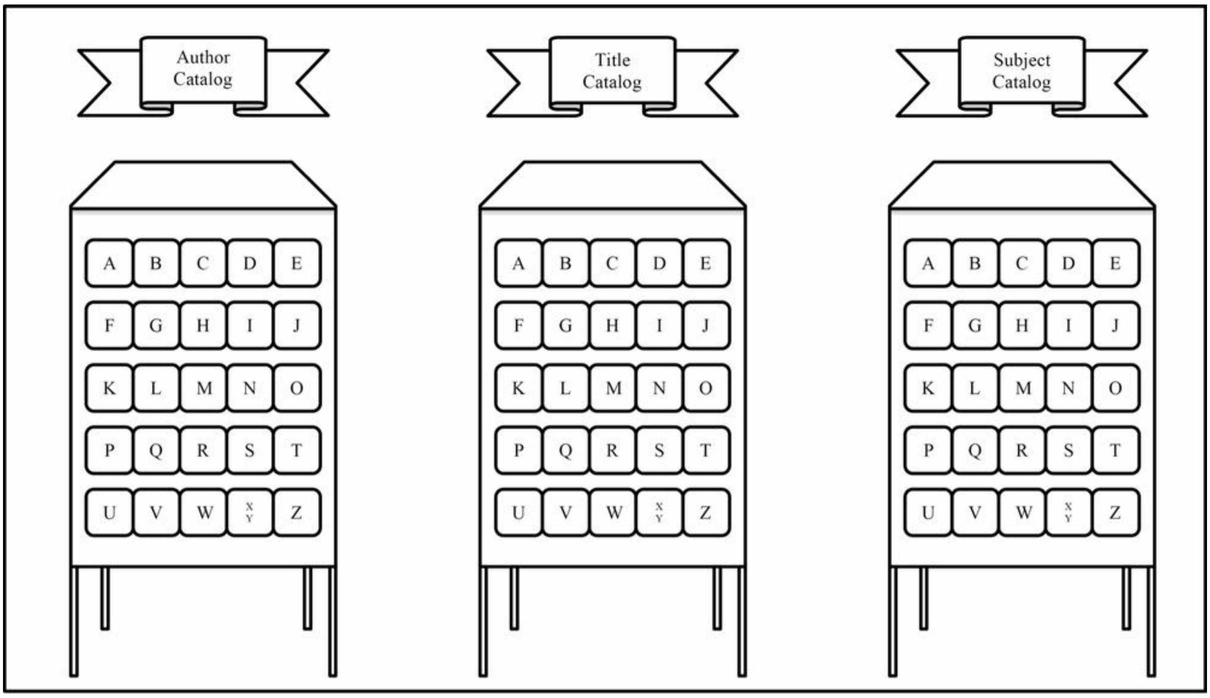


Figure 3.8. A Divided Alphabetical Card Catalog.

With the development of OPACs in the 1980s, the divided catalog was moved online. One usually had to search either by author or by title or by subject (although, later, more options for searching became available). In this case, though, it was seldom possible to retrieve works both by and about a person in the same search; this was not possible until general keyword searching became available in the online catalog, and even then it was not terribly precise because the name might appear in parts of the record other than a creator or subject field. General keyword searches are similar to the dictionary catalog approach to arranging a catalog, because one does not have to choose the type of search one wishes to conduct; search terms are retrieved no matter where they appear in the catalog record. The classified arrangement is also available in today's OPAC; it is usually achieved by performing a call number search, *if* that type of search is available (which is not always the case). In some ways, the sophistication that had been built into the design of card catalogs through elaborate intellectual filing rules was abandoned in exchange for the convenience of automated searching.

Other Arrangements. Displays of search results today are not always in alphabetical or classified order. The specific OPAC's programming determines which types of arrangements of search results are possible and when each arrangement is used as the default display. In some catalogs, searches for creators or titles may bring up an alphabetical display of authority-controlled access points for creators or a list of titles in alphabetical order, but other systems have a default setting that is used for almost every type of search: *relevancy ranking*. Arrangement by relevance can be mysterious to users and information professionals alike. The idea is that the resources that are most closely connected to the search terms are ranked higher in the search results. Relevancy is determined algorithmically, and it is defined differently by each system. It is based on a combination of factors (e.g., placement, field weighting, term frequency) that may be adjusted somewhat according to an institution's preferences.

The choice of display arrangements and the choice of searches available vary from catalog to catalog. Some, for example, do not provide users with the ability to conduct a classification or call number search, but they may allow title or creator search results to be sorted in call number order. Some OPACs do not allow browsing of authority-controlled lists of names, instead focusing primarily on keyword searching. Most online catalogs now allow users a choice of the order in which search results are to be displayed. Catalogs may allow sorting by

- relevance ranking,
- creator's name (alphabetical by last name),
- title (alphabetical by first non-article word),
- call number,
- ascending chronological order (publication date from earliest to latest),
- descending chronological order (from latest to earliest),
- most popular resources first, or
- format or material type.

The choices offered by catalogs may vary from just two or three of these sorting options to, perhaps, many or most of the types listed. In some systems, these arrangements can also include options for a variety of sub-arrangements under the primary sort key (e.g., sorting first by creator but then choosing sub-arrangement by title or publication date).

Indexes

Indexes are retrieval tools that provide access to the analyzed contents of information resources (e.g., articles in journals, short stories in collections, papers in conference proceedings). Although *back-of-the-book indexes* provide access to the analyzed contents of one work, they are not retrieval tools in the sense defined in this chapter; they are prepared at the time of publication of the work, not at a later time in an effort to provide bibliographic control. They do, however, aid with retrieval of the information found in the text at hand. Some websites also have indexes that function very much like back-of-the-book indexes, but instead of page locator references, each entry links directly to an HTML anchor within the site being indexed. Online these are called *A–Z Indexes* (to distinguish them from search trees and search engines; see discussion below). A–Z indexing employs specialized software (such as HTML Indexer and XRefHT32).[26] Like back-of-the-book indexes, an A–Z web index should contain properly identified second-level entries, variant entries, and cross-references.

Indexes that are retrieval tools in the sense discussed in this chapter are separate from the information resources being analyzed. *Database indexes* (also called *journal indexes* or *periodical indexes*) are the longest-lived examples of indexes as retrieval tools. Such indexes often do not have authority control of names. A name such as Lois Mai Chan can be entered into the same index at different times as Chan, L.; Chan, LM; Chan, L. M.; Chan, Lois M; Chan, Lois Mai; and even Mai Chan, Lois. They do, however, usually maintain authority control over subjects. Indexers tend to use hierarchical controlled vocabularies (called *thesauri*) for the topical terms they wish to bring out in the index. A thesaurus contains systematized language that subsumes narrower terms under broader terms and provides a structure of relationships between related terms. (See Chapter 10 for more information on thesauri and other controlled vocabularies.)

Unlike catalogs, indexes are not limited to what is available in a local setting or a particular collection, and they do not usually give location information as such. They do give, as part of the description, the larger resource in which the smaller work can be found. Especially with print indexes, it may then be necessary for a user of the index to search a catalog for the location of the larger work. In some online indexes, there exists the capacity to link to online catalogs to allow users to see quickly if their library owns the larger work. If the larger work is not found in the local catalog, the user may have to search a union catalog or make a request through interlibrary loan (ILL). Increasingly, online indexes have links to full text versions of articles.

Indexing can be carried out by people or by machines or by a combination of both. Database indexing is still mostly completed by specially trained indexers, although they may have some machine assistance. For many decades, indexes were available in print form, but today they are primarily online tools. Some print versions were arranged in alphabetical dictionary fashion, with entries for authors, titles, and subjects interfiled. Others reflected a divided arrangement, having an author/title part separate from the subject index. Online indexes have interfaces that dictate how they are searched and how results are displayed. As with OPACs, there is no standardization from index to index. OPACs, at least, have the standardization that comes from using RDA (or another set of descriptive rules) whereas there is no such standard commonly used by every index. The National Information Standards Organization (NISO) tried to update Z39.4–1984, "Basic Criteria for Indexes,"[27] but committees could not come to agreement with the American Society for Indexing, and so Z39.4 was withdrawn in 1996. A standard from the International Organization for Standardization (ISO), which the British and others have adopted, was published in 1996.[28]

Another difference between catalogs and indexes is that indexes tend to be created by for-profit organizations, such as H. W. Wilson, or professional societies, such as the American Chemical Society. Often there is a charge for access to the online versions. Libraries pay for the right to allow their patrons to use the indexes online without cost. Some libraries have both print and online access, although there seems to be a trend to continue only with the online versions. In any case, the index often is cataloged so that the whole index can be found through the catalog.

The sample index entry in Figure 3.9 illustrates the types of information included in an index for scholarly journal articles. It begins with the title of the article and then lists various characteristics of the article, including who wrote it, contact information, and affiliations. There is also information about the journal, including its title, type, and an identifier, usually the International Standard Serial Number (ISSN). Other information includes the publisher's name, the volume and issue in which the article is found, the page numbers, and the article's length. There is also an indicative (descriptive) abstract and associated subjects.

Most of the indexing already discussed is performed by humans. However, as early as the 1950s people began working on automatic indexing that would rely on the power of computers to perform repetitive tasks at high speed. The first such method, introduced by Hans Peter Luhn in 1958, became known as KWIC (*Key Word In Context*) indexing. In a KWIC index, titles are arranged so that each of the words in a title appears once in alphabetical order in the center of a page, with all other words to the left or right of the center word printed in the order in which they appear in the title (so that the alphabetized word is *in context*).

Describing Entities and Identities: The Development and Structure of Encoded Archival Context—Corporate Bodies, Persons, and Families.

Author:	**Wisser, Katherine M**.
Email Address:	wisser@simmons.edu
Author Affiliations:	School of Library and Information Science, Simmons College, USA
Source:	**Journal of Library Metadata**
Source Type:	Scholarly journal
ISSN:	1938-6389
Volume:	11
Issue:	3-4
Date:	Jul-Dec 2011
Pages:	166-175
Length:	10 pages
Publisher:	Taylor & Francis, Philadelphia, PA
Document Type:	Journal Article
Subjects:	Archives Encoded Archival Context-Corporate Bodies, Persons, and Families International Standard Archival Authority Record for Corporate Bodies, Persons & Families
Abstract:	In January 2011, the Society of American Archivists fully endorsed the standard Encoded Archival Context-Corporate Bodies, Persons, and Families (EAC-CPF). This article details the development of the standard, from its conceptual beginnings in 1998 to its dissemination and adoption in 2011. It provides an overview of the general structure of EAC-CPF and discusses variables that were considered important in the design of the standard. It concludes with a reflection on the strength of international participation in the development and review of the standard in order to ensure that EAC-CPF would be applicable across many different boundaries. [ABSTRACT FROM AUTHOR]
DOI:	10.1080/19386389.2011.629960
Accession Number:	12349876
...	

Figure 3.9. A Representation of an Index Entry for a Scholarly Journal Article.

KWIC

Introduction to	**Cataloging** and Classification.
Introduction to Cataloging and	**Classification**.
	Introduction to Cataloging and Classification.

An adaptation of the KWIC method is known as KWOC (*Key Word Out of Context*), which prints the significant words in the left-hand margin instead of in the middle of the page, and the title is printed to the right or beneath the keyword.

KWOC

Cataloging	Introduction to Cataloging and Classification.
Classification	Introduction to Cataloging and Classification.
Introduction	Introduction to Cataloging and Classification.

Two additional types both identified as *KWAC* are used somewhat less frequently. The first, *Key Word And Context* is a combination of KWIC and KWOC, in which the significant word appears in both the heading position and in context with the rest of the title. The second, *Key Word Augmented in Context*, allows the indexer to augment the title with additional words to provide more meaning and to enhance access to the resource.

KWAC: Key Word And Context

Cataloging and Classification. Introduction to
Classification. Introduction to Cataloging and
Introduction to Cataloging and Classification.

KWAC: Key Word Augmented in Context

Cataloging and Classification [in Libraries]. Introduction to
Classification [in Libraries]. Introduction to Cataloging and
Introduction to Cataloging and Classification [in Libraries].
Libraries. Introduction to Cataloging and Classification

The current application of automatic indexing is in search engine indexing. A search engine is sometimes referred to as an index but, more accurately, a search engine creates lists of terms that are referred to as indexes. A *search engine* is a computer program that searches web documents for specified keywords via a program called a *spider*. The list of documents is returned to an internal database and placed in order by the indexing program according to the words or concepts contained in the document. With the development of microdata and similar annotating schemes, general markup vocabularies (such as Schema.org), and the use of other encoded metadata tags, search engine indexing *could* become more accurate for users. Search engines are described more fully below.

Finding Aids

Finding aids (sometimes called *inventories*) are descriptions of archival collections, and they tend to be longer than typical catalog records. Archives usually maintain control over entire collections (not individual pieces) of archival materials from personal, corporate, or institutional sources. Thus a finding aid describes a collection, not a single letter or record within that collection. While some finding aids are not publicly accessible, this is becoming increasingly rare as archives add more finding aids to the web. The finding aid, itself a retrieval tool, is often linked to a summary description of the collection in a catalog record. Summary records provide name, title, and subject access points for archival collections in the catalog and also provide a means to point users to more detailed finding aids. *Describing Archives: A Content Standard* (DACS) is used in the United States to create both finding aids and MARC-based catalog records. An example of a brief finding aid is presented in Figure 3.10.

A finding aid may be a brief or quite lengthy description, depending upon the size of the archival collection, the complexity of its organization, and the level of granularity desired in the description. Because the materials in archives are different from the types of information resources found in other institutions, the types of metadata recorded in a finding aid are considerably different from that found in catalogs or indexes. The following types of information are found in a typical finding aid:

- **Provenance—**including creator, title, dates, and bio-graphical or administrative history
- **Physical extent and condition—**including the amount (number, volume, and linear or cubic feet) and a statement of the collection's condition
- **Scope and content notes—**including a narrative descrip-tion of the nature of the collection, what is included, the extent and depth of the collection, gaps in the collection, types of materials, and access points for names and subjects

Guide to the Industrial School for Girls (Dorchester, Boston, Mass.) records, 1873-1934

Archivist's View

Descriptive Summary

Creator	Industrial School for Girls (Dorchester, Boston, Mass.)
Title	Industrial School for Girls (Dorchester, Boston, Mass.) records
Dates	1873-1934
Identification	CC 30
Quantity	0.5 linear feet (1 manuscript box)
Collection Abstract	The Industrial School for Girls collection contains the annual reports of the Board of Managers, arranged chronologically with a concentration of reports from 1873 to 1934. Also, there is a paper read at the 50th anniversary of the Dorchester Industrial School on June 7, 1904, as well as a bound copy of *Suggestions to Visitors of Dependent Children,* 1874.
Historical Abstract	The Industrial School for Girls was established in 1853 and incorporated 1855 in the names of Lucretia O. Everett and Maria Greenwood. It moved from Winchester, Massachusetts, to Dorchester, Massachusetts, in 1858, located at 232 Centre Street. It provided a home and training school for various branches of housework to develop the habits and principles to become upright, self-supported women. By the mid-1940s the Industrial School for Girls had evolved into the Everett House located at the same place, and in the 1950s the New England Home for Little Wanderers acquired it.
Language	Material in English.
Location	Collection may be stored offsite. Please contact Archives staff for more information.

Information for Users

Access Restrictions

Collection is open.

Copyright Notice

Copyright for materials resides with the creators of the items in question, unless otherwise designated.

Publishing Permission

Please contact the College Archivist with requests to publish any material from the collection.

Preferred Citation

[Identification of item: description and date], Industrial School for Girls (Dorchester, Boston, Mass.) records, CC 30, Simmons College Archives, Boston, MA, USA.

Acquisitions Information

Transferred from the Simmons College School of Social Work Library, 1991

Accession number: 2002.178

Processing Information

Processed by Molly Tierney, December 2002; Supervised by Jason Wood. This collection guide was encoded as part of the LEADS project by Katie Sallade, November 2012

Organizational History

The Industrial School for Girls was established in 1853 and incorporated 1855 in the names of Lucretia O. Everett and Maria Greenwood. It moved from Winchester, Massachusetts, to Dorchester, Massachusetts, in 1858, located at 232 Centre Street. It provided a home and training school for various branches of housework to develop the habits and principles to become upright, self-supported women. They accepted on average 25 students through an application process, ... girls between 10 to 14 years of age, whose family or friends are unable or unfit to care for them. The original age of acceptance was between the ages of 6 to 10, which gradually rose with modern social standards. Reductions of boarding charges were made when relatives could not pay. The girls attend public schools (this started in 1881) and Congregational church. They would go out to earn their living as soon as able under the immediate care of the head of household, usually in country families, and each one, unless returned to relatives, would be supervised under the care of one of the managers. By the mid-1940s the Industrial School for Girls had evolved into the Everett House located at the same place. And in the 1950s the New England Home for Little Wanderers acquired it.

Information taken from *Directory of Charitable and Beneficent Organizations*, Boston, 1907 and 1940 or the *Report of the Board of Managers*, 1926 and the finding aid for CC 6, "Guide to the New England Home for Little Wanderers records."

Collection Overview

The Industrial School for Girls records consist of the annual reports of the Board of Managers, arranged chronologically with a concentration of reports from 1873 to 1934. Also, there is a paper read at the 50th anniversary of the Dorchester Industrial School June 7, 1904, which has a summation of the first 50 years of the School and a bound copy of Suggestions to Visitors of Dependent Children, 1874, which provides guidance for guardians of the girls.

There has been an annual report printed every year except in 1858 when they were moving from Winchester to Dorchester. The reports cover officers' positions, brief history of the School, student population, admission application statistics, requests for student help placement, expense reports, list of subscribers for previous year, donations made and by-laws of the organization. Missing are the years: 1919, 1923, 1925, 1928, 1930, and 1932.

Online Catalog Headings

These and related materials may be found under the following headings in online catalogs.

Charities -- Massachusetts -- Boston
Dorchester (Boston, Mass.)
Industrial School for Girls (Dorchester, Boston, Mass.)
New England Home for Little Wanderers -- History
Winchester (Mass.)

Collection Arrangement

Collection is arranged into 3 series:

Series I: Annual Reports
Series II: Manual
Series III: 50th Anniversary Paper

Related Material

Part of the School of Social Work Library Charities Collection.

Detailed Description of the Collection

Series I: Annual Reports, 1873-1934 (3 folders)

This series contains the Industrial School for Girls' annual reports printed each year except in 1858. The reports cover officers' positions, brief history of the School, student population, admission application statistics, requests for student help placement, expense reports, list of subscribers for previous years, donations made, and by-laws of the organization. Missing years include: 1919, 1923, 1925, 1928, 1930, and 1932.

Box 1

- Folder 1: **Bound 1873-1885, 1886-1900**
- Folder 2: **1874-1920**
- Folder 3: **1921-1934**

Series II: Manual, 1879 (1 folder)

This series contains a bound copy of *Suggestions to Visitors of Dependent Children*, 1874, which provides guidance for guardians of the girls.

- Folder 4: **1879**

Series III: 50th Anniversary Paper, 1904 (1 folder)

This series contains a paper read at the 50th anniversary of the first 50 years of the Dorchester Industrial School on June 7, 1904, which has a summation of the first 50 years of the School.

- Folder 5: **1904**

Figure 3.10. An Example of a Finding Aid. (Source: Beatley Library Collection: CC30, Simmons College Archives, Boston, MA, USA)

- **Order and structure—**including a description of how the collection is arranged and organized, and detailed container lists
- **Administrative information—**including location, conditions of use, restrictions, processing, citation, acquisition and accession, repository, and conservation information

Registers and Other Museum Databases

Registers constitute the primary collections control tool for museums. A register may also be called an *accession log.* It functions like a catalog, although it has additional kinds of access points. The process of registration in a museum is much like the process of cataloging in a library. During the process, the registrar identifies the object, the donor, any associations (e.g., having belonged to a particular person), any information needed for insurance purposes, and so forth. An identification number is assigned. This process of documentation is internal and seldom made available to the public. The accession record becomes the basis for one or more files that help provide organization and control of the museum's content.

Additionally, museums sometimes provide databases for public access online. For example, the Pitt Rivers Museum in Oxford, England, has an objects database and a photos database. One can search each database by filling in fields in a search record. The objects database (also called the *Objects Collection Catalogue*) has search boxes for such things as name of the object, country the object is from (a pull-down menu provides the current United Nations list of country names), cultural group from which the object comes, material made from, when collected, and so forth. In addition to the same fields as for objects, the photo database (also called the *Photographs Collection Catalogue*) has search boxes for such things as activity depicted, event depicted, person shown, photographer, photographic process, image dimensions, date of photograph, and so forth. It is suggested that users should not try to search more than three separate fields in the database at one time.[29] This approach is very factual, with a good deal of information provided by the databases, but it lacks a stylized presentation of the collection that may be found at larger museums.

An example of a more comprehensive and stylized approach is that found at the National Gallery of Art (NGA) in Washington, DC. The entire collection (drawings, paintings, photographs, prints, and sculptures) is searchable in at least two distinct methods through their website. The first interface is marked "Search the Collection." It allows the public to search the collection's "object records" by artist name, keywords in title or in the object description, donors and owners, accession number, exhibition history, and a few other fields.[30] After an initial search is complete, users may filter their search results by medium, nationality, time span, styles, and whether an object is on exhibit. A digital image for many, but not all, works is provided. In addition, some artworks have extensive descriptions. For example, the search results for Edward Hopper's 1939 painting *Cape Cod Evening* begin with some brief information:

- **Material/Technique**: oil on canvas
- **Dimensions**:
  - **Overall**: 76.2 × 101.6 cm (30 x 40 in.)
  - **Framed**: 106.7 × 132.1 cm (42 x 52 in.)
- **Collection**: John Hay Whitney Collection
- **Accession number**: 1982.76.6

Following this, there is, among other things, an essay that describes the history of the work along with bibliographic references, images of initial sketches used to create the painting, a technical summary and notes, descriptions of the work's provenance, its exhibition history, and resources where the work has appeared.[31] Some artworks in the collection have significantly more information provided than others.

The other interface available is through the NGA Images repository. This separate retrieval tool allows users to search by keyword anywhere in the description of the artwork or in the title, by artist's name, and accession number. It provides search filters for nationality, classification (the type of art: painting, sculpture, media art), medium, and dates. The search results provide downloadable digital images for many of the works, a link to the object record (mentioned above), and a brief description:

- **Artist** Luis Meléndez
- **Artist Info** Spanish, 1715–1780
- **Title** Still Life with Figs and Bread
- **Dated** c. 1770
- **Medium** oil on canvas
- **Classification** Painting
- **Dimensions**
  - **Overall**: 47.6 × 34 cm (18 3/4 × 13 3/8 in.)
  - **Framed**: 65.6 × 49.2 × 4.4 cm (25 13/16 × 19 3/8 × 1 3/4 in.)
- **Credit** Patrons' Permanent Fund
- **Accession No.** 2000.6.1
- **Digitization** Direct Digital Capture

- **Image Use** open access[32]

In addition to searching for particular artworks, the National Gallery of Art also provides a searchable index of artist's names, which can be filtered by nationality, dates, the first letter of the last name, and so on. This is one particular museum approach to providing some information about their collections to the general public. Other cultural heritage institutions approach it in other ways. As time goes along, we may or may not see more consistency among presentations of retrieval tools on museum websites.

Search Engines

For the majority of information seekers, search engines are probably the most familiar of all the retrieval tools discussed in this chapter. Search engines are tools developed for computer systems, particularly the Internet, to find instances of requested words or phrases that can be found in the documents covered by the scope of the tool. They were developed for the purpose of searching full text documents (or indexes of those documents) for particular words or phrases. Some search engines are used to search only a single particular resource, such as the website for an organization such as OCLC. Other search engines are for searching the entire Internet (although only a small percentage of the Internet is actually searched by each one). Using a program called a *spider* (also known as a *crawler* or *robot*), each search engine automatically collects information from web resources and places it into a database of records or in a full-text index that is similar to a concordance. The spider is typically programmed to go onto the web to retrieve and download copies of web pages and everything linked to them, everything linked to the links, and so on. Massive collective indexes—created from the spider's database—are stored in data centers around the world. These indexes (some comprising over 100 million gigabytes of data) are what are searched when users enter terms into the search box.

Obviously, not every web resource has been indexed. Some resources may be indexed by only a few search engines—thus, a variety of responses can be had when searching different engines for the same concept. In addition, not every search engine indexes every type of material available online. It has only been in more recent years that resources in PDF and PowerPoint format have become searchable online. In some cases, resources may be password protected, and as a result, they may not be indexed in a search engine because they are not accessible to the crawler. In other cases, some resources may be indexed and returned as search results, despite being restricted—a fact that the searcher may discover only after trying to access the information. In general, only the surface web is included in search engine indexes; the so-called *deep web* (or hidden web) includes many things the general public would not want to see indexed in search engines (e.g., email correspondence, financial transactions, dating profiles, nefarious dealings of all kinds).

Unlike catalogs and finding aids, search engines have become a ubiquitous part of users' routine interactions with the world (at least for those advantaged enough to have consistent and reliable Internet access). Due to a general familiarity with search engines and their relative ease of use (just throw some keywords into the search box!), other retrieval tools often look antiquated and complex in comparison. Although search engines may or may not provide results that are as intellectually satisfactory as the results from other retrieval tools, users often report satisfaction because they find *something* related to what they were searching for, and find it fast. But most users do not know if what they found is authentic, authoritative, or the best that is available on their topic. Searches for known items or for specific names are sometimes less satisfactory than searches for topical information. Some users do not always realize that search engines may push advertisements and sponsored links to the top of the results page, or that more in-depth information may be found much further down in the list of results.

There is no doubt that search engines have become more advanced in recent years. Google[33] is one of the most sophisticated both in giving searching assistance (e.g., asking, "Did you really mean to search for . . .?" in response to a misspelled word) and in display of results, but there are still drawbacks. At the time of this writing, there is still no real distinction among homographs. For example, on the first page of search results for the term *bridge*, Google brings back a mixture of resources about card games, structures spanning geographic spaces or gaps, specific named highway structures, and some local organizations with the word *bridge* in their names. The handling of synonyms and related terms is somewhat mysterious. In Google, searching *funny pets* brings back videos about funny pets, but the top search results on the page are for *funny dogs* (where the word *pets* is nowhere to be found on the page). Although humans understand that dogs are a type of pet, machines do not and cannot know this; they have to be programmed to connect these strings of characters and to treat them as related concepts. Although Google acknowledges that it has a "synonyms system" of which they are very proud and that it involves analyzing "petabytes of web documents and historical search data to build an intricate understanding of what words can mean in different contexts,"[34] Google is not particularly forthcoming about how synonym control works. Disappointingly, its synonyms system is not something a user can explore directly.

With most search engines, displays of search results are usually arranged according to *relevance,* which can be calculated in various ways. A search engine may calculate relevance by giving different weightings to factors such as how many search terms are found in each web page, how often each term is found in a page, whether the terms are in proximity or dispersed, whether the terms are in the head of the document or buried further down in its body, and how common (or rare) a particular search term is. Search engines continue to strive for more sophisticated methods of calculating relevance and for displaying results. For example, Google uses a formula that examines over 200 different factors in determining relevance. Google uses common factors, such as word placement and word frequency, in its formula, but also less-expected criteria such as site quality, geographic region, search history, recency, and site popularity (i.e., how often a page is linked to by other web pages) as elements in the ranking of search results.[35] Google, Bing, and other search engines are continuously updating, refining, and testing their algorithms and formulas in their quests to improve the Internet searching experience.

When metadata was first being developed to describe web documents, most search engines did not take user-supplied metadata into consideration in their relevance ranking due to metadata mischief and malfeasance (e.g., including popular terms that have no relationship to the site being described). With the development of linked data, however, the major search engines are now embracing the inclusion of more structured data as part of the underlying textual markup of web resources. Encoding approaches such as JSON-LD (Java Script Object Notation for Linked Data), microdata, and RDFa (Resource Description Framework in Attributes)—in combination with specialized vocabularies (such as the categories established in Schema.org) to identify necessary semantics—may be used to add more contextual metadata directly into web resources. For example, when searching recipes on a culinary website, it would be very helpful if the search engine could distinguish between dietary restrictions associated with the recipe and a recipe's ingredients. With support for linked data and for processing contextual metadata found encoded within web documents, search engines are expected to continue to improve over the coming decades.

Directories

In the 1990s and early 2000s, there were still some alternatives for finding web resources other than through the use of search engines; these alternatives were *directories* and *subject gateways*. As discussed in Chapter 1, directories were organized collections of links to websites. The directory created by Yahoo![36] was one of the largest and the best known. In its heyday, Yahoo! employed human indexers to create and maintain its directory. Humans added subject terms and categorized the records into a hierarchical subject tree index. Searching was done by browsing (or "drilling down") through the categories or by keyword searching. A small number of Internet directories were organized according to traditional library classification, but most had organization schemes that were created by the persons who devised the directory, usually with 12 to 16 top-level categories (e.g., Arts, Business, Computers, Finance, Games).

Directories and subject gateways, once the first place to look for websites, have waned considerably in popularity due to the ascendancy of the search engine and also to the tremendous efforts and resources needed to maintain extensive lists of ever-changing web resources. Most of the long-standing Internet directories and subject gateways have ceased operations, but there are a few that remain open. To get a sense of how these directories were organized, and what searching for information on the web was like in the previous millennium, please visit one of the last large-scale Internet directories: the Best of the Web directory.[37]

THE NEED FOR RETRIEVAL TOOLS REVISITED

In today's heavily web-focused information environment, it is often tempting to think that nothing more than a search engine is ever needed. Some feel that if a resource cannot be found through Google's web, scholar, or book search, then that resource may not be worth finding. This, however, could not be further from the truth. It can be easy to forget that much of the world's information is *not* online at this time and that much of that information probably never will be. In addition, a great deal of what *is* available in electronic form is not necessarily *freely* available to all. Information is still a commodity. Most e-books are not free; they are sold. A needed article may be behind a paywall. There can be costs associated with information access online.

Although this may change sometime in the future, individual retrieval tools (beyond search engines) still have vital roles to play in information gathering. Catalogs, indexes, and other retrieval tools, however, are often compared negatively to search engines. Students and scholars alike have said, "Why is the catalog still hard to use?"[38] and "Why can't the catalog be more like Google [or Amazon.com or Facebook]?" The desire for straightforward and user-friendly tools is understandable and pervasive, and there *are* information professionals who are working to rectify the situation. However, we must be careful to remember that this is not a contest. Not everything is as simple as putting keywords into a single Google search box. There are times that users and information professionals need more advanced types of searching that are currently not available in the search engine approach (e.g., an author search rather than a general keyword search).

Different types of retrieval tools have been designed to describe different kinds of materials for various communities. Information professionals must understand where and how to access the best information for each particular situation. For example, one must access a finding aid if one wants detailed information about an archival collection—relying on the brief catalog record for the collection usually will not suffice. Using Google to get to the Wikipedia page for Eleanor Roosevelt might be a fine starting point for a school report, but it will not provide comprehensive information about her life and accomplishments; published print materials, such as biographies and histories, will provide a greater depth of information—and a catalog will easily retrieve those materials (as long as they are in that collection). As psychologist Abraham Maslow said, "I suppose it is tempting, if the only tool you have is a hammer, to treat everything as if it were a nail."[39] We must be very cautious about overreliance on a single type of retrieval tool. We must not let the web search engine become our hammer. The trick is to make sure that all the tools in the toolbox are understood, and that they are used when necessary and appropriate.

CONCLUSION

A century ago Charles Cutter stated that catalogs should enable people to find books for which titles, authors, or subjects were known; should show what was available in a library by a particular author, on a particular subject, or in a kind of literature; and should assist in the choice of a book by its edition or character. A century later, Cutter's "objects" are still quite appropriate, except that they have been expanded to all kinds of information resources beyond just books in libraries and to a number of different kinds of retrieval tools beyond catalogs.

This chapter has discussed the major retrieval tools used in the organization and retrieval of recorded information. The surrogate records that make up a retrieval tool must be created and encoded, either by humans or automatically by software programs created by humans. The individual retrieval tools discussed in this chapter are types of information systems. In the next chapter, we provide an overview of some of the issues and problems in organizing information as they relate to information systems and system design. Chapters on metadata and encoding then follow, before settling into a multi-chapter discussion of the descriptive metadata found in retrieval tools.

Some Important Terms in This Chapter

A–Z Index	Directory	Printed Catalog
Access Point	Divided Catalog	Register
Accession Log	Explore	Relevance
Alphabetical Catalog	Find	Relevancy Ranking
Alphabetico-classed Catalog	Finding Aid	Research Guide
Annotation	Identify	Retrieval Tool
Authority Control	Index	Robot
Bibliography	Inventory	Search Engine
Catalog	Journal Index	Select
Citation	Known-Item Searching	Shelf Reading
Classified Catalog	Library Guide	Spider
Computer Output Microform	Metadata	Subject Gateway
Crawler	Microdata	Subject Guide
Database Index	Obtain	Surrogate Record
Deep Web	Online Public Access Catalog	Thesauri
Description	Pathfinder	Union Catalog
Dictionary Catalog	Periodical Index	

Some Important Acronyms in This Chapter

COM:	Computer Output Microform
FRBR:	*Functional Requirements for Bibliographic Records*
IFLA:	International Federation of Library Associations and Institutions
IFLA LRM:	*IFLA Library Reference Model*
JSON-LD:	Java Script Object Notation for Linked Data
OPAC:	Online Public Access Catalog
RDFa:	Resource Description Framework in Attributes
UBC:	Universal Bibliographic Control

NOTES

All URLs accessed June 2017.

1. *Publication Manual of the American Psychological Association,* 6th ed. (Washington, DC: American Psychological Association, 2010).

2. *Chicago Manual of Style,* 16th ed. (Chicago: University of Chicago Press, 2010).

3. *MLA Handbook,* 8th ed. (New York: Modern Language Association of America, 2016).

4. Council of Science Editors, *Scientific Style and Format: The CSE Manual for Authors, Editors, and Publishers,* 8th ed. (Chicago: Council of Science Editors in cooperation with the University of Chicago Press, 2014).

5. Kate L. Turabian, *A Manual for Writers of Research Papers, Theses, and Dissertations,* 8th ed., revised by Wayne C. Booth, Gregory Colomb, Joseph M. Williams, and the University of Chicago Press editorial staff (Chicago: University of Chicago Press, 2013).

6. U.S. Government Printing Office, *Style Manual,* 30th ed. (Washington, DC: U.S. Government Printing Office, 2008). Also available at: https://www.gpo.gov/fdsys/pkg/GPO-STYLEMANUAL-2008/pdf/GPO-STYLEMANUAL-2008.pdf.

7. Kirsten Madden, Michèle Pujol, and Janet Seiz, *A Bibliography of Female Economic Thought Up to 1940* (London: Routledge, 2012).

8. David Gilson, *A Bibliography of Jane Austen* (Winchester, UK: St. Paul's Bibliographies, 1997).

9. Mustafa Aydin and M. Nail Alkan, *An Extensive Bibliography of Studies in English, German, and French on Turkish Foreign Policy, 1923–1997* (Ankara, Turkey: Ministry of Foreign Affairs, Center for Strategic Research, 1997).

10. John T. Zepper, *Russian and Soviet Education 1731–1989: A Multilingual Annotated Bibliography* (Hoboken, NJ: Taylor & Francis, 2014).

11. Richard T. Wang, *Area Bibliography of China* (Lanham, MD: Scarecrow, 1997).

12. *Faculty Bibliography*, prepared by the reference staff of the University of Iowa Law Library (Iowa City: University of Iowa Law Library, 2015).

13. Boston Public Library, *The Stinehour Press: Work of the First Fifty Years* (Boston: Boston Public Library, 2003).

14. John McIlwaine, *Maps and Mapping of Africa: A Resource Guide* (London: Zell, 1997).

15. For an example of a high-quality subject guide, please take a look at http://libguides.library.albany.edu/iberoamericanonlinevideo.

16. LibGuides, http://www.springshare.com/libguides/.

17. SubjectsPlus, http://www.subjectsplus.com/.

18. Charles A. Cutter, *Rules for a Printed Dictionary Catalog* (Washington, DC: Government Printing Office, 1876), 10.

19. Seymour Lubetzky, *Cataloging Rules and Principles: A Critique of the A.L.A. Rules for Entry and a Proposed Design for Their Revision* (Washington, DC: Processing Dept., Library of Congress, 1953), 36.

20. Ibid.

21. International Conference on Cataloguing Principles, Paris, 9th–18th October, 1961, *Report* (London: International Federation of Library Associations, 1963), 91–96. Also available in: *Library Resources & Technical Services* 6 (1962): 162–67; *Statement of Principles Adopted at the International Conference on Cataloguing Principles, Paris, October, 1961,* annotated ed., with commentary and examples by Eva Verona (London: IFLA Committee on Cataloguing, 1971); and A. H. Chaplin and Dorothy Anderson, eds., *Report/International Conference on Cataloguing Principles, Paris, 9th–18th October 1961* (London: IFLA International Office for UBC, 1981).

22. International Federation of Library Associations and Institutions, IFLA Study Group, *Functional Requirements for Bibliographic Records* [FRBR]: *Final Report* (Munich: Saur, 1998), 82, https://www.ifla.org/publications/functional-requirements-for-bibliographic-records.

23. See, for example, Arlene G. Taylor, "Cataloging in Context," in Bohdan S. Wynar and Arlene G. Taylor, *Introduction to Cataloging and Classification,* 8th ed. (Englewood, CO: Libraries Unlimited, 1992), 8.

24. Pat Riva, Patrick Le Boeuf, and Maja Žumer, "IFLA Library Reference Model: A Conceptual Model for Bibliographic Information," August 2017 (The Hague: IFLA, 2017), 15–16, https://www.ifla.org/publications/node/11412.

25. Martha M. Yee and Sara Shatford Layne, *Improving Online Public Access Catalogs* (Chicago: ALA, 1998). For more about standardization in online systems, as well as more about OPACs, see Chapter 5.

26. David M. Brown, "Introducing HTML Indexer," Brown, Inc. Available at: http://www.html-indexer.com; Tim Craven, "Tim Craven—Freeware: XRefHT32," University of Western Ontario. Available at: http://publish.uwo.ca/~craven/freeware.htm.

27. *American National Standard for Library and Information Sciences and Related Publishing Practices—Basic Criteria for Indexes, ANSI Z39.4–1984* (New York: American National Standards Institute, 1984).

28. International Organization for Standardization, *Information and Documentation: Guidelines for the Content, Organization and Presentation of Indexes,* ISO 999:1996 (Geneva: ISO, 1996).

29. Pitt Rivers Museum, "Online Databases and Catalogues," http://www.prm.ox.ac.uk/databases.html.

30. National Gallery of Art, "Search the Collection," http://www.nga.gov/content/ngaweb/Collection/collection-search.html.

31. Robert Torchia, "Edward Hopper/Cape Cod Evening/1939," *American Paintings, 1900–1945,* NGA Online Editions, http://purl.org/nga/collection/artobject/61252.

32. National Gallery of Art, "NGA Images: Advanced Search," https://images.nga.gov/en/search/show_advanced_search_page/?form_name=defaultname. Perform an artist's name search for *Meléndez, Luis* to see the painting and the record.

33. Google, http://www.google.com.

34. Steven Baker, "Helping Computers Understand Language," *Official Google Blog,* January 19, 2010, https://googleblog.blogspot.com/2010/01/helping-computers-understand-language.html.

35. Google, "How Search Works," https://www.google.com/search/howsearchworks/.

36. Yahoo!, http://www.yahoo.com.

37. Best of the Web Directory, https://botw.org/.

38. Christine L. Borgman, "Why Are Online Catalogs *Still* Hard to Use?" *Journal of the American Society for Information Science* 47, no. 7 (1996): 493–503.

39. Abraham H. Maslow, *The Psychology of Science: A Reconnaissance* (New York: Harper & Row, 1966), 15.

SUGGESTED READINGS

Bettington, Jackie, ed. *Keeping Archives*. 3rd ed. Canberra: Australia Society of Archivists, 2008. Chapter 12: "Finding Aids."

Breeding, Marshall. *Next-gen Library Catalogs*. New York: Neal-Schuman, 2010.

Buckland, Michael K. "What Is a 'Document'?" *Journal of the American Society for Information Science* 48, no. 9 (September 1997): 804–9. http://people.ischool.berkeley.edu/~buckland/whatdoc.html.

Cleveland, Donald, and Ana Cleveland. *Introduction to Indexing and Abstracting.* 4th ed. Santa Barbara, CA: Libraries Unlimited, 2013. Chapter 1: "Introduction," Chapter 2: "The Nature of Information," Chapter 3: "The Organization of Information," and Chapter 8: "Types of Indexes."

Joudrey, Daniel N., Arlene G. Taylor, and David P. Miller. *Introduction to Cataloging and Classification.* 11th ed. Santa Barbara, CA: Libraries Unlimited, 2015. Chapter 1: "Cataloging in Context."

Kumar, Shiv. "From Clay Tablets to Web: Journey of Library Catalogue." *DESIDOC Journal of Library & Information Technology* 33, no. 1 (January 2013): 45–54.

Library of Congress. "Manuscript Division Finding Aids Online." https://www.loc.gov/rr/mss/f-aids/mssfa.html.

Markey, Karen. "The Online Library Catalog: Paradise Lost and Paradise Regained?" *D-Lib Magazine* 13, no. 1/2 (2007). http://www.dlib.org/dlib/january07/markey/01markey.html.

Miller, Laura. *Archives: Principles and Practices*. New York: Neal-Schuman, 2010.

OCLC. *Online Catalogs: What Users and Librarians Want: An OCLC Report*. Dublin, OH: OCLC, 2009.

Reibel, Daniel B. *Registration Methods for the Small Museum*. Lanham, MD: AltaMira, 2008. Chapter 6: "The Catalogue."

Roe, Kathleen. *Arranging & Describing Archives & Manuscripts*. Chicago: Society of American Archivists, 2005. Chapter 4: "The Practice of Arrangement and Description."

Rowley, Jennifer, and Richard Hartley. *Organizing Knowledge: An Introduction to Managing Access to Information.* 4th ed. Burlington, VT: Ashgate, 2008. Chapter 9: "Organizing Knowledge in the Digital Environment."

Salo, Dorothea. "The Humble Index." *The Book of Trogool*. Blog posting August 25, 2009. http://scientopia.org/blogs/bookoftrogool/2009/08/25/the-humble-index/.

CHAPTER 4

SYSTEMS AND SYSTEM DESIGN

This chapter provides an overview of some of the issues and problems in organizing information as they relate to information systems and system design. In order to ensure resource discovery, systems must be designed to store, retrieve, and display metadata in useful and logical ways. The retrieval tools described in the previous chapter require an adequate system design to allow users to find information they seek. This chapter looks at system components that assist professionals to organize information and assist users to retrieve needed information resources; it also addresses innovations in system design and a range of other systems issues related to the information organization process. Some system design issues, many technical in nature, are not addressed here because they go beyond the scope of this book.

SYSTEMS

The *Dictionary of Computing* defines the term *system* as "anything we choose to regard (a) as a whole and (b) as comprising a set of related components. . . . In computing the word is freely used to refer to all kinds of combinations of hardware, software, data and other information, procedures, and human activities."[1] Therefore, the management of systems may involve a single source of data handled through a single type of software or a more complex mixture of various data sources, pieces of hardware, and software components.

In information organization, the term *system* generally refers to what we call an *information system* or an *information retrieval system.* An information system performs three basic functions: (1) *storage* (data organization), (2) *retrieval* (based on queries), and (3) *display* (the interface or presentation design). Each one of these functions is necessary and vital to the system, and the operation of each depends on the other two. System design decisions take into consideration all three of these functions. Data is stored in a system using a specific format that is suitable for containing the type of information being amassed. The data stored in the system is partitioned into logical groupings (e.g., in tables or indexes) based on the nature or function of that data. The groupings are based on how the individual pieces of data are encoded or tagged. For example, all of the creator metadata in a system is stored together in one index, and this metadata is identified by particular codes or tags such as <creator>, <contributor>, <agent>, <author>, the 100 and 700 MAchine-Readable Cataloging (MARC) field tags, and so on. To access the information, a user or another system formulates a search string, which is converted to the query language used in the system in order to retrieve the data stored there. Quick retrieval is achieved through the indexing process. It allows the query to look at only the appropriate index(es) rather than trying to match the query to individual pieces of data in individual records, one record at a time. Once the data is retrieved, the proper presentation renders the raw data usable and understandable.

Among the institutions discussed in this book, libraries were the first to automate. Though the question "Why automate?" is rarely asked these days, Larry Milsap reminds us that libraries automated in order to

- provide access to the complete catalog from multiple locations,
- increase and improve access points,
- increase and improve search capabilities,
- eliminate or reduce inconsistencies and inaccuracies of card catalogs,
- reduce the increasing problems and costs of maintaining card catalogs, and

- deal with pressures and influences for change.[2]

With the move to automation came major changes in the ways libraries performed daily tasks and fulfilled their obligations to patrons. Archives, museums, and other cultural heritage institutions also embraced the use of information technology in making their information resources accessible to users. Today, the automation revolution is a thing of the past; the importance and prevalence of information technology in organizing processes and in institutions are just assumed. From yesterday's quills and typewriters to today's mobile apps and linked data structures, technology has been and will continue to be a vital part of information organization. This does not mean, however, that all of the problems of systems and organizing information have been resolved. Information professionals still have much work to do, and it is unlikely that all the problems will be completely resolved before new questions come along.

Databases

One of the most basic tools used for organizing information is the *database.* According to *FOLDOC: Free On-line Dictionary of Computing,* a database is "one or more large structured sets of persistent data, usually associated with software to update and query the data."[3] In other words, databases are organized collections of data. They provide the structure that underlies many of our information systems. Typically, a database is a set of records or data elements, each representing a specific entity, all constructed in the same way (with common attributes), and connected by relationships. The records may contain numeric information, text, or graphic representations.

Records are the basic components of a database. A record comprises data fields or elements (such as creator's name, stock number, employee's date of birth, etc.), which describe chosen attributes of an entity (a book, a product, an employee, or any other entity). A database *can* be represented in a paper format, but it is nearly always thought of as machine-readable data. Databases become necessary when there is too much information for humans to process and analyze by themselves. An everyday example to help clarify the basic components of a database is a checkbook, or an online banking activity log. Each transaction has a record (e.g., a row in the checkbook). Transactions have attributes, such as date, check number, payee, transaction description, amount, and adjusted balance. In each column, there is a field for holding the value of each attribute. The attribute values identify, describe, and explain the transactions.[4]

Date	**Check #**	**Description/Payee**	**Amount**	**Balance**
1/17/16	**132**	**American Library Association Student Membership Fee**	**$ 36.00**	**$ 495.34**

Of course, most databases are more complex than this, but the checkbook example does contain all elements of the basic database structure. It should be noted that a variety of organizational methods may be used to manage the data stored inside a database, and these methods vary from system to system. A key feature of bibliographic databases is the use of indexes created to hold different types of information. For example, separate indexes may be created to track creator data, title data, subject data, and the like; the number and types of indexes that are created vary from system to system.

Today, most database applications are *relational databases.* These are databases designed using an entity-relationship model. In the relational database model, the collection of metadata typically found in a surrogate record is divided into parts (entities), which are held in various tables. These parts are linked to each other to form individual metadata displays that show how the various entities are related. Each individual piece of information is stored in only one place, reducing data redundancy, but it may be used in multiple displays. For example, an author's name may be stored only once in a table of names, but the display for each work written by that author includes the author's name. To interact with the information held in a database, a query language is used. *Structured Query Language* (SQL) is the common programming language used to manage and retrieve information contained within most relational databases.

In the past, some systems were based on *hierarchical databases,* which used a traditional tree structure as the model for holding information. They consisted of one file composed of many records, which in turn were made up of numerous data fields. These databases tended to be rather inflexible and used more space as data was often repeated. Hierarchical structures are rarely used in today's information systems, and the ones that exist are being replaced primarily by relational models. Another option is the use of native *Extensible Markup Language (XML) databases,* which process the data and tagging of that data in the same files. In the early 2010s a growing number of developers and users began developing various types of alternative database structures lumped under the heading *non-relational databases*—sometimes called *NoSQL* databases (*NoSQL* stands for *not only SQL*). Non-relational databases have a number of sub-types including graph databases, document databases, object-oriented databases, and more.[5] The more technical aspects of database structures are beyond the scope of this work, but what is important to understand is that more than just a single type of database is found in the larger computing environment.

Databases serve various functions. They may hold administrative data, a collection of images, or raw numerical data. They may contain surrogate records or hold the actual information resources of interest. They may be repositories of full-text articles or they might keep track of inventory and sales. The functions of databases can be divided into two categories: *reference databases* and *source databases.* Reference databases contain links or pointers to information sources held outside of the database, for example, a journal index containing information about the location and contents of articles that are stored elsewhere. Source databases contain the actual information sought, for example, a human resources database containing employee information. Databases created as information retrieval tools generally contain surrogate records. The retrieval tools described in the previous chapter can all be held in computer databases, although bibliographies and finding aids are more likely to be displayed as online documents. Databases underlie almost all the tools that we use to organize information. Indexing and abstracting services, book sellers, museums, and libraries, to name a few, all rely on databases to hold the records of their inventories.

Bibliographic Networks

Bibliographic networks have as their main resource a huge database of catalog-type records. Access to the database is available for a price, and members of the network can contribute new records and download existing ones. Bibliographic networks acquire many MARC records from the Library of Congress (LC) and other subscription sources. The databases also include cataloging records contributed by participating libraries. In either case, the records contain two kinds of information: (1) cataloging data (such as title, subject, creator, dates), and (2) holdings information for libraries that have added specific items to their collections.

Bibliographic networks were organized in the 1970s to support library technical services operations through cooperative cataloging and computer-assisted card production. Although they have steadily expanded their activities, their continued emphasis on cooperative cataloging most clearly distinguishes them from other online information services that provide access to similar surrogate records. For example, LC MARC records are available online through other information services, but these do not offer online entry of original cataloging data, record editing, or other services that specifically support cataloging services.

The three major bibliographic networks discussed here are OCLC, SkyRiver, and Auto-Graphics. Bibliographic networks differ in their administrative structures and customer bases. OCLC Online Computer Library Center (formerly known as the Ohio Computer Library Center) is the largest and most comprehensive bibliographic network, which operates as an international, nonprofit membership organization.[6] Auto-Graphics (formerly known as A-G Canada, and before that as ISM/LIS, which grew out of UTLAS at the University of Toronto) operates for-profit, as does SkyRiver. All three are general-purpose bibliographic networks that are available to libraries of all types and sizes. Another bibliographic network, Research Libraries Information Network (RLIN), was specifically created to meet the special needs of research institutions, but in 2006 it merged with OCLC and is no longer a separate entity. OCLC offers services to libraries in 120 countries with a database of more than 402 million records in over 491 languages. More than 16,000 libraries or other institutions worldwide are OCLC members.[7] OCLC encourages participation through regional networks that act as its authorized agents. Customers may access these bibliographic network databases through web interfaces.

SkyRiver contains more than 60 million full, unique records including the entire set of LC MARC records; like WorldCat, it grows through records contributed from member libraries.[8] Auto-Graphics is used mostly in Canada, with some customers in the United States; it offers MARCit, a cooperatively developed database consisting of over 30 million bibliographic and authority records, including the LC bibliographic database and user-contributed records.[9] Although it is typical for organizations to subscribe to only a single bibliographic network, some libraries in the past used both RLIN and OCLC for different purposes or types of materials.

Bibliographic network databases are specifically designed to automate cataloging and allow for record sharing. If existing cataloging copy is available, it is displayed in the full MARC format, or in OCLC it can alternatively be displayed in Dublin Core format. A record editor can be used to move, change, delete, or otherwise edit individual field values to meet local cataloging requirements (called *copy cataloging*). If no cataloging copy is available, a cataloger may create a new record (called *original cataloging*). To facilitate cataloging decisions and promote consistency, bibliographic networks provide online access to the LC name and subject authority files. The OCLC bibliographic network has an interlibrary loan (ILL) subsystem, which is an obvious use of a union catalog, and it offers a variety of *retrospective conversion* products as well (in which paper-based catalog records are converted to machine-readable form).

Integrated Library Systems (ILSs)

Another type of system in use in libraries is the *integrated library system* (ILS). The ILS is more than just an *online public access catalog* (OPAC); it is a fully integrated data management system that includes multiple modules to perform different functions. Its purpose is to incorporate the functionality of five or six separate database tools into one integrated system, all using the same database. The greatest benefit comes from information sharing among the various modules, reducing duplication of data and effort. An ILS may have modules to support the following:

- acquisitions
- authority control
- cataloging
- circulation
- course reserves
- digital object management
- interlibrary loan
- public access (the OPAC)
- serials management
- system administration

This system may also be known as an *integrated library management system* (ILMS), *library management support system* (LMSS), or sometimes as a *library housekeeping system* (LHS), particularly in the United Kingdom.

More recently, additional terms have augmented the profession's lexicon to describe a new generation of library systems. These include *webscale management solution*, *uniform management system*, and most commonly, *library services platform* (LSP). These are systems that are intended to go beyond the capabilities of the traditional ILS. Carl Grant explains that the primary difference is that the ILS has been designed to manage print collections, but an LSP integrates seamlessly the management of both print and digital resources (which are often quite different processes).[10] Marshall Breeding, who coined the term *library services platform*,[11] lists some of the characteristics of an LSP. An LSP

- enables libraries to acquire and manage their collections, spanning multiple formats of content;
- supports multiple acquisitions processes, including those for ownership, paid licenses and subscriptions, and open-access sources;
- supports multiple metadata schemas, including *at a minimum*, MARC and Dublin Core;
- may include an integrated discovery service or support a separately acquired discovery interface;
- provides all staff and patron functionality though browser-based interfaces; and
- provides *knowledgebases* that represent the body of content extending beyond the library's specific collection.[12]

An LSP also incorporates more current technologies and architectures, such as cloud computing, which are often unavailable in the traditional ILS.[13]

History of Library Systems

Before the first ILS appeared on the scene, computers were already being used to organize information. As mentioned in the previous chapter, computers were used in catalog production in the 1960s and 1970s to produce book and microform catalogs. At that time, computers were not the desktop machines we think of today but were instead large mainframe systems that were run in batch mode. A quick search in the computer was not possible; jobs (e.g., queries) were sent in batches and often the results were not available until the next day. Record formats were defined locally, as the MARC standard was still in its infancy. It was not until the late 1970s that minicomputers, which could provide some online access to information, began appearing in libraries.[14] In order to have online catalogs, there had to be a convergence of computing power, mass storage, low costs, software that could handle large files, and the files themselves in electronic format.[15]

The first automated library systems were created in the 1970s. Many of these commercially produced products were *turnkey systems* (computer systems customized to include all the hardware and software necessary for a particular function or application). Often these products were developed to handle only one specific type of function, such as circulation. Then as more vendors got into the library automation business (increasing competition), additional modules were created to expand system functionality. Numerous systems failed, but by the 1980s a second generation of online catalogs began to appear with increased search capabilities.[16] In recent years, some smaller and midsized vendors have survived by focusing on specific types of libraries or by catering to specific niches.[17] A number of early systems were created as in-house products, but by the end of the twentieth century, most of those systems had been replaced by commercially developed packages. LC was one of the last holdouts before it announced in 1999 that it would replace its 30-year-old in-house computer systems with the commercially produced Voyager system. A major reason that LC replaced its homegrown approach (which consisted of seven separate record systems) was that data in an ILS can be shared among the different modules.

Developments in Library Systems

In the last 30 years, there have been tremendous improvements in integrated library systems. Although many of these developments have centered around the OPAC, others have affected the entire system. Some of the major developments included the creation of *Web-PACs* (web-accessible catalogs), the move to graphical user interfaces (GUI), widespread implementation of client-server architecture, the use of industry-standard relational databases, support for Unicode to enable the use of multiple scripts and multiple languages, implementation of self-service features, personalized interfaces, and the creation of authentication mechanisms for remote users.[18]

Over the years, library automation vendors have continued to innovate. In recent years, they have developed new approaches to digital object management, collection management that integrates print and electronic resources, license and rights management, and search and discovery. Of these, the innovations in resource discovery tools (or *discovery interfaces*) have received the most attention. Breeding states:

> Reacting to the need for libraries to compete with other web destinations, much product development industry energy focused on development and marketing of new web-based interfaces. The battle is on to deliver interfaces that showcase faceted browsing, relevance-ranked results, end user rating or tagging, and visual navigation—all standard fare in commercial web interfaces.[19]

Some of these resource discovery tools were created by companies outside of the ILS vendor marketplace (such as Endeca[20] or AquaBrowser[21]), while others were created by the vendors themselves (such as Encore by Innovative Interfaces[22] and Primo and Summon by Ex Libris[23]). Most of these products profess to work with *all* major ILS/LSP systems and to use XML encoding to combine information from MARC records, user tags, web-based data, and other sources in one seamless interface.

In recent years, a growing interest in open source software (OSS) has led to the development of a few viable open-source systems for libraries, such as the Koha ILS,[24] the Evergreen ILS,[25] and the VuFind discovery interface.[26] Compared to the commercial products in the ILS marketplace, the number of open source systems is still rather small, but OSS ILSs may gain greater market share in the future. Breeding states, "The success of early adopters' implementations has already diminished skepticism. Many indicators suggest that open source ILS contracts will displace larger percentages of traditional licensing models in each subsequent year."[27] Both Koha and Evergreen "are well established among specific library sectors."[28]

In general, the ILS/LSP marketplace continues to shrink. In 1990 there were approximately 40 companies providing library automation software solutions; in 2017, because of a large number of mergers and buyouts, only about half are still in business. For more information about current trends in library systems, see Breeding's annual "Library Systems Report" in *American Libraries*.[29]

Development of Online Public Access Catalogs

The OPAC is the most familiar module in a library system, and it is the one most in need of system design attention. Two decades ago, Charles Hildreth stated, "The online public access catalog is the first major development that brings the benefits of automation directly to the user as a means of expanded access to library collections and as a means of organizing and presenting bibliographic information for effective self-services."[30] The OPAC has been a focus of system design research for many years. System design issues are examined below after looking at the developmental stages of the OPAC.

The evolution of the OPAC is often described in terms of *generations.* The first generation of OPACs appeared in the early 1980s as crude finding lists, often based on circulation system records[31] or simple MARC records, perhaps with a circulation, serials, or acquisitions module added on.[32] Based on card catalogs and early online retrieval systems, their searching capabilities were limited generally to authors and titles, using only *left-anchored searching* (i.e., all searches must be based on the first word or text string starting at the left). For example, in left-anchored searching the title *Introduction to Cataloging and Classification* must be searched starting with the word *introduction,* and it will not be found by a search on the word *classification.* The interface of first-generation OPACs was menu-based and fairly primitive. These early systems had no subject access or reference structures. They were little more than poor imitations of print catalogs. Some systems were programmed to respond to commands in which a code (e.g., *a:* for author, *t:* for title) was to be followed by an exact string of characters that would be matched against the system's internal index. In others, derived key searching was supported (i.e., taking parts of names and/or titles to create a search string; for example, *int,to,ca,a* would be used to search for *Introduction to Cataloging and Classification*).[33] These first-generation systems were highly intolerant of user mistakes. There was little or no browsing and little or no keyword searching with or without Boolean operators (i.e., *AND*, *OR*, *NOT*). Access points were limited to those that had been available in the card catalog. First-generation OPACs were primarily book finding lists and worked best for known-item searching.[34] In addition, in many early systems the display of search results was by the "last in, first out" principle (i.e., the last records entered into the system [or pulled out, reworked, and reentered] were those listed first in the display).

With designers learning from the problems of the first generation, the second generation of OPACs in the late 1980s showed major improvements. This generation was marked by significantly improved user interfaces. Keyword searching, with its use of Boolean operators, was introduced, thus increasing the number of access points available for searching. This meant that searches were no longer required to be exact left-anchored word or phrase searches; words could now be matched, even if they were in the middle of a text string. Also greatly enhancing the searching process were truncation and wildcard support, browsing capabilities (including index term browsing), use of full MARC records, interactive search refinement, and subject access to items. Second-generation OPACs also provided greater manipulation of search results and provided better help systems with more informative error messages (although there is still a lot of work to be done in this area).[35]

Through the first and second generations of OPACs, the characteristics distinguishing each generation are fairly clear. As we move beyond the second generation, however, there are differences in how the profession refers to the more recent developments in OPACs. Some consider the systems that are currently in use (WebPACs with GUIs, Z39.50 compliant systems, etc.) to be third-generation OPACs. Others describe the third generation as catalogs that are still in experimental stages. Hildreth acknowledged the improvements made in catalogs during the 1990s but referred to these improved catalogs as *E^3 OPACs* rather than as third generation. The E^3 OPAC (a phrase that never really caught on) was so named because it was *enhanced* in functionality and usability; *expanded* in indexing, data records, and collection coverage; and *extended* through linkages and networks, acting as a gateway to additional collections. It was marked by an improved, more intuitive GUI. Hildreth believed additional improvements would be made over time.[36]

In the first decade of the twenty-first century, discussions shifted away from how the current OPACs can be improved to boisterous discussions about how current OPACs were inadequate (actually the language used was a bit more colorful than that) and how changes were needed.[37] The question on everyone's mind was how to design a *next generation catalog.* These discussions, which were taking place on various blogs, forums, conference meetings, and discussion lists were mostly focused on determining just what the future OPAC should be and what it should do in this increasingly web-based, "Google-philic" information environment. Some suggestions for next generation catalogs included the following:

- **Simpler interfaces** (similar to Amazon.com or Google): This has been embraced in today's library system. Many of today's search systems—ILSs, LSPs, and discovery interfaces—begin with a single search box. In fact, it can be difficult in some tools to locate browsing functions or advanced searching.
- **Expanded searching**: Through an LSP or a discovery layer, many libraries provide their users the ability to search all parts of the collection at the same time. No longer do searchers have to search the catalog for print materials and search separate databases for electronic full-text articles; both may be searched with a single, integrated search box, such as that found in the EBSCO Discovery Service.[38]
- **Increased direct access**: In both discovery interfaces and catalogs, access to more than just surrogate records is provided. In OPACs, direct links to full-text documents and digital objects may be provided. In discovery layers, full-text articles may be easily accessed through a single interface, as long as the searcher is an authenticated user and a link resolver is in place.
- **Faceted browsing**: This feature has also come to pass, mostly in discovery layers that allow users to browse keyword search results in new ways. Browsing can be done using facets such as subject, genre, format, call number, library branch, language, geographic area, time period, and so on. An example of faceted browsing can be found in the Minuteman Library Network's Encore-enhanced union catalog.[39]
- **Increased interaction**: In recent years, catalogs and discovery interfaces have allowed for more user input, similar to Web 2.0 applications where users review, rank, recommend, or tag information resources. These types of activities can be performed in Library Thing[40] (a web tool for organizing collections of resources) and in catalogs that use VuFind as their discovery interface (for example, the library catalog at the University of Illinois).[41]

These advances represent an exciting area where information organization and information technology intersect. While many of the changes hoped for have come to pass, we are still in a period of transition, and it is expected that lively discussions of where we are heading will continue for the foreseeable future, particularly as linked data principles become more influential in the management and dissemination of our data. Linked data functionalities are being tested in many systems. In addition, libraries are facilitating access to information that has been structured through linked data platforms such as the United States government data.gov site.[42] Stay tuned for further developments.

SYSTEM DESIGN

Organization of Information and System Design

At times, it is unclear where the process of organizing information ends and where system design begins. In a display of metadata records from a retrieval tool, both aspects come together to present to the user the information sought. This set of results combines features of the organizing process (standard punctuation, forms of access points, etc.) and features of system design (labels, screen layout, etc.) in the presentation of the metadata. When the metadata is clear, understandable, easily retrieved, and well presented, the user usually does not notice system design or organizing elements. It is only when there are problems or confusion that these elements are discussed.

In the print world, system design was not separate from the process of creating surrogate records. Panizzi's rules included principles for what information to include in surrogate records and also standards for placing those records into a cohesive catalog. Cutter's rules included record-creation rules that emphasized collocation (i.e., placing those records in logical juxtaposition with one another) and also included a section of filing rules (i.e., Cutter's design for the card catalog). Each edition of cataloging rules published in the twentieth century by the American Library Association has assumed the system design of a card catalog; although with the publication of *RDA: Resource Description & Access* in 2010, this is no longer the case. Standards for the creation of bibliographies and indexes often assume a print format, generally in book form.

Of course, the term *system design* was not used for the process of deciding how print tools would be arranged and laid out. The same people who created surrogate records also controlled the display of those records. They did not think of themselves as system designers. Yet some print tools were and still are quite sophisticated finding and collocation systems. They could be sophisticated because the designers knew the contents of the surrogate records intimately. As the tools became automated, though, the task of design was taken on by people who understood computers but often had little or no knowledge of the contents of the records that would make up the system. System design is a necessity for the retrieval of organized information, whether it is or is not done by the same people who create surrogate records. It can be a design that simply displays, in no particular order, every record that contains a certain keyword. Or it can be a design that displays records in a sophisticated way to show relationships among records and among works, as well as responding to requested search words.

System design research can be divided into two categories: technology-focused research and user-centered research. These two categories are not mutually exclusive. There are overlapping concerns and interconnections between the two (e.g., the system's search functionality and the user's information-seeking behavior have important connections). Users' needs and search behavior *should* influence the design of the technological systems we use (whether they do is another issue). While this book is not an appropriate forum for a treatise on information retrieval and information-seeking behavior, it is helpful to start with the basic underlying assumption that users approach retrieval tools with an information need. How they meet that need is greatly affected by the characteristics of the system they encounter.

Searching Methods

According to Hildreth, the two basic approaches to searching, *querying* and *browsing,* can be subdivided into more specific methods. Querying can be *phrase matching* or *keyword matching.*[43] Phrase matching is matching a particular search string to the exact text located in records in the system (or more precisely, to particular indexes created by the system). This type of query demands that the words of the string be found together in the same order as given in the search query (i.e., left-anchored searching). If, for example, users were searching for works by Lev Grossman, a left-anchored phrase-matching search would work well. The user would simply enter the following string into the search box:

Creator: Grossman, Lev

The system would then search the creator index (only), and search the name letter-by-letter.

G-R-O-S-S-M-A-N L-E-V

Phrase matching does not allow for the terms or strings to be found in various fields (or indexes). Doing this type of search, you would not retrieve *Grossman* from the creator index and *lev* from the contents notes (as in *mag-lev trains*).

Allowing the terms to be separated is a characteristic of the second type of query: keyword matching. Keyword searching involves matching discrete words to the system's indexes, often using Boolean operators or proximity formulations to combine them. Keywords may be matched against terms that occur in more than one field or index. For example, a general keyword search for *Twelfth Night* will bring back records for Shakespeare's play, but it will also bring back, among many other things, a volume of essays titled *Hemingway's Spain*, because the essays come from the *Twelfth* Biennial International Hemingway Conference and one is about Hemingway's story, "*Night* Before Battle."

Browsing, too, can be divided into two sub-categories. Pre-sequenced, linear, system-based browsing allows users to scan lists of terms, names, or brief titles to find topics, creators, or items of interest. This is the more structured approach, using the system's internal organization of the data to guide browsing activities. The second type of browsing is nonlinear and multidirectional. This is browsing that is more serendipitous. It is unstructured. It uses or can use hypertext links to navigate between various items and may be more exploratory or seemingly more random. In recent years, the phrase *faceted browsing* has come into use; it refers to system-based browsing that exploits various types of data found in metadata records. This type of browsing is discussed further below.

While querying is useful when the users know exactly what they want (i.e., known-item searching), browsing in many cases may be a preferable alternative to retrieving hundreds of records in response to a query. For example, if a user is looking for an item by an author with the last name *Shaw,* but only has an initial *L* for the forename, a browsing approach will put the searcher into an index of all personal names with the surname *Shaw* at the point where the L-forenames would begin. From that point, the user may scan names either before or after the entry point into the list. If this listing contains, in addition to the name, the entire authoritative access point for the author, including dates and the number of records associated with that name, then the user can more easily determine which name represents the *L Shaw* sought. A subject browsing option would provide the user with a list of subjects that surround the word that was input as the search. In cases where the user is unsure of what is wanted, or when the user does not know the exact string used in the system to describe what is sought, browsing may provide a more manageable approach to meeting the user's information need.

Retrieval Models

Despite the presence of browsing in retrieval tools, our current systems are mostly oriented toward exact-match queries, in which the exact specifications of the query must be satisfied by the document representation in order to create a match. In order for a system to work fairly well, users must have a good idea of what they are looking for. This is an all-or-nothing approach. It is precise and rigid, purely mechanistic, and the burden is placed on users, not on the system. If there is no exact match, then nothing is retrieved.

However, other models do exist. For example, probabilistic retrieval is based on term weighting (which is based on word frequency). Probabilistic systems return results that match the query to some degree and are displayed in the order of decreasing similarity. Such models have been the subjects of research for many years. They represent attempts to find ways around the limitations of exact-match, Boolean-based retrieval. These retrieval models are not without criticism. They do not take into account information-seeking behaviors such as browsing or exploratory searching, which are not query-based. In the best of all possible systems, a variety of search options and retrieval techniques would be available.

Over the years, many researchers and practitioners have made suggestions for improving online retrieval systems, but the creators of these systems have seemingly not paid attention to many of the suggestions. Evidence of this comes from two articles by Christine Borgman. In 1986 she wrote an article titled, "Why Are Online Catalogs Hard to Use?"[44] A decade later she wrote another article, "Why Are Online Catalogs *Still* Hard to Use?"[45] Borgman stated that OPACs were hard to use because their design still did not reflect an understanding of user searching behavior, which still rings true decades later. She suggested that a long-term goal of system designers should be to design intuitive systems that require a minimum of instruction. Instruction has often been touted as the way around inadequate system design, but in this age of remote access, where instruction may not be available, one has to agree with Borgman: "Good training is not a substitute for good system design."[46]

Standardization and Systems

David Thomas mentions that a user's need for standardized description has not been acknowledged in online systems, leading to a loss of familiarity for the user.[47] Standardization was a key feature of the card catalog. Users who were familiar with one catalog could apply the same knowledge and skills to other catalogs they encountered with a minimum of difficulty. Standardization, however, does not happen overnight. It takes time to develop. The lack of standardization today is reminiscent of more than a century ago, when catalogs contained cards of varying sizes (e.g., 2 × 5 inches, 2 × 10 inches, 3 × 5 inches) and the information placed on cards lacked a standard order. Standardization came when Melvil Dewey's 7.5 × 12.5 centimeter catalog cabinets won out over other sizes.[48] In the library cataloging community, throughout the latter part of the twentieth century, there were calls for standard interfaces for online catalogs. As Martha Yee states, "The lack of standardization across OPACs can make it difficult for catalogue users to apply their knowledge of one OPAC to searching another OPAC in a different library."[49]

Today, there are still no real standards, just broad guidelines and suggestions. Due to the competitive nature of vendors in the library systems marketplace, a standard interface is unlikely to occur anytime soon. Vendors develop new features, and each vendor places different levels of importance on various aspects of its system design. Some have diverse internal organization schemes; some have discrete search capabilities. Vendors try to develop the most appealing interface, search features, and modules in order to gain a greater market share. This competition contributes to the lack of standardization in system design, but it can also contribute to long-term system innovations and progress. This may be a question of finding the right balance between standardization and market forces. Online indexes, too, have a great deal of variety and little standardization. This is a continuation of a long history of lack of standardization in print indexes. It seems that commercial enterprises have little interest in standardizing.

Some areas in which standardization has been recommended include the following:

- Display
- Basic search queries
- Treatment of initial articles
- Use of Boolean operators, proximity, and truncation
- Punctuation

These are discussed below.

Display

One of the key areas in which the lack of standardization is most apparent is in system displays. Displays can be divided into two categories: (1) the display of sets of retrieved results and (2) the display of metadata in surrogate records. Both incorporate issues of screen layout and design.

Display of Retrieved Results. The first concern in the display issue is whether the initial search results appear as a list of individual records or as a list of access points displayed first before the actual records are presented for viewing. An example may help clarify the problem here. Some systems, in response to a search for an author with the surname *Benson,* display all Bensons in the system, alphabetized by forename. One can browse through this list and find the appropriate access point for the desired Benson before having lists of resources to sift through.

74 **Benson, Sally, 1897–1972.**
75 **Benson, Sidney William, 1918–**
76 **Benson, Sonia.**
77 **Benson, Stephen, 1969–**
78 **Benson, Susan Porter, 1943–2005.**
79 **Benson, Thomas W.**
80 **Benson, Warren, 1924–**

Other systems return results that are lists of works related to each Benson without identifying first how many Bensons there are.

1 Love and money / by Mr. Benson.
2 The Amen sisters / Angela Benson.
3 This is the end / Stella Benson.
4 The last dream keeper / Amber Benson.
5 A million blessings / Angela Benson.
6 U.S. immigration and migration almanac / Sonia Benson.
7 Beading for the first time / Ann Benson.
8 Miami / Sara Benson.
9 Meet me in St. Louis / Sally Benson.
10 The improvisation of Musical Dialogue / Bruce Benson.

One is required to page through the list of titles—which may or may not be in an order that is self-evident—instead of being able to browse the list of creator names first.

The second concern in displaying results is the order in which results are shown. As mentioned earlier in this chapter, first-generation catalogs often worked on the “last in, first out” principle. As systems matured, they began to display responses to specific searches, such as those for author or title, in alphabetical order sorted by a certain field. In catalogs and indexes, the principal or first-named author was usually chosen as the field by which responses would be arranged. However, results of less specific searches, such as keyword, were sometimes displayed in reverse chronological order—using date of publication or record creation—or in order of relevance as defined by that particular system. As systems have become more sophisticated, more control of the display has been given to the user. In most systems, users can specify, from a short predefined list of options, how they wish search results to be sorted. However, such sorting may still present problems. For example, if one chooses to have results sorted by author, then the access point for the principal or first-named author is displayed in the sorted list, but that access point may not be the author that the user knows about or is interested in (e.g., the Benson being sought could be a second or later author of a resource).

In recent years, some institutions have added a discovery interface to their information system, which allows *faceted browsing*. Faceted browsing relies on the metadata indexes created by the system in order to provide new and interesting ways of grouping search results. Instead of simply putting a keyword into a search box and retrieving a large randomly returned set of results, faceted browsing systems cluster the results based on certain aspects of the data (i.e., facets), such as subject matter, format, genre, language, classification number, region, time period, and so on. By clustering the results according to these facets, users have the ability to narrow the set to only the most pertinent results without having to scroll through page after page of results ranked by system-defined relevance.

Display of Records. When dealing with the display of individual records several questions must be addressed. These include:

- Does the display provide an appropriate amount of information?
- Which fields should be displayed?
- What labels should be used?
- Will patrons be able to view the information on a single screen?
- In what order should the information appear?

The type of information contained in catalog records is guided by long-established and sometimes elaborate rules. However, in local situations many records may be either longer (e.g., addition of contents notes, reviews) or shorter (e.g., minimal records) than the common standard. Similarly, information contained in index records is quite variable. Each commercial index has its own standard for inclusion of information in records. Differences may include length of citation, abbreviations, inclusion of abstracts, and the presence or absence of controlled vocabulary terms.

In the case of MARC records, some data that a cataloger includes in a record often goes unseen by the public (e.g., many pieces of coded information). In a good system design, coded information could be programmed to be displayed or to be searched in meaningful ways, but neither of these options is universally available. Commercial indexes use their own encoding schemes, perhaps MARC-based or possibly internally created. They display records in a variety of ways, often allowing the user some control over the types and amount of information they see. Other kinds of metadata have no typical ways of being displayed.

There are often multiple levels of display in an online catalog. There may be a one- or two-line version, a brief display, and a full display. Although *full* seldom means that all information in a record is displayed, some systems do allow the display of an entire MARC record, complete with MARC tagging, to the public. The default display, when only one record is retrieved in response to a query or after a user has selected a record to view from a list, is often a brief view. The default display, however, is different from system to system. The amount and kinds of information omitted from a full record to make a brief display also differ from system to system. Allyson Carlyle and Traci Timmons surveyed 122 web-based catalogs and determined that personal author, title, and publication fields are almost always displayed in default single records, but other fields are displayed less frequently, and the ones that are displayed are treated inconsistently (e.g., title fields sometimes include statements of responsibility, but sometimes do not).[50] Thomas notes that users find only a small number of fields useful; so there should be guidelines for selecting the most-needed fields for display.[51] Once employed, these guidelines could also help to reduce screen clutter.

The labeling of metadata in records also varies from system to system, and there are differences in the terminology used for labeling. Often, records for serial publications (i.e., magazines, journals, newspapers, and other periodicals) suffer from labeling problems because of the complexity of information found in those records. Holdings information (e.g., metadata about the issues and volumes of a periodical to which a library can provide access) can be confusing for users even with the most explicit labels available. In addition, the need for labeling metadata at all has been questioned. Labels can be confusing and do not necessarily cover everything in a field. For example, in the MARC format, the 245 field, usually labeled *title,* also contains a statement of responsibility (i.e., the names of those responsible for the information resource). Thomas points out that although many assert the need for labels, there is no empirical evidence that labeled displays are more effective.[52] One could assume, then, that using unlabeled displays would cause few or no problems for users, and the need for brief records could be eliminated. However, labels and brief displays persist.

One difficulty with brief displays is that what is included in them is inconsistent from system to system. Some displays include only the first note, while others suppress all notes. Notes are designed to give information that clarifies data given in other areas of the description. Without those notes, users may be misled by the information that is given. Other data can also be suppressed in brief displays. In some systems, subject headings are not presented. Even more problematic, some show no statements of responsibility or additional access points. If a user has searched for a secondary creator or a contributor to the resource, the brief display may not mention that person, family, or corporate body being sought. These display limitations can cause users to overlook potentially relevant or useful resources based on what is initially shown.

In addition to differences in record displays, a user is often faced with major differences in the order of lists of retrieved results. If the list is chronological, it may be in ascending order or in descending order. If the list is supposed to be alphabetical, there are numerous ways a computer can be programmed for alphabetical order, affected by punctuation, spacing, and other such factors. Numerical order is likewise unpredictable; not to mention the problem of whether numbers come before or after letters. These and other issues are discussed in Appendix C: Arrangement of Metadata Displays.

Display Guidelines. These different approaches to both types of displays (i.e., display of retrieved results and display of records) have led some to work toward guidelines for standardizing displays. A task force of the International Federation of Library Associations and Institutions (IFLA) in 1999 issued guidelines to assist libraries in designing or redesigning their OPACs. The guidelines consist of 37 principles based on the objectives of the catalog and the types of searches that users conduct.

> The guidelines recommend a standard set of display defaults, defined as features that should be provided for users who have not selected other options, including users who want to begin searching right away without much instruction. . . . The goal for the display defaults recommended is ease of use, defined as the provision of as much power as possible with as little training as possible. If such defaults were widely implemented, users would benefit by being able to transfer catalogue use skills acquired in one library to many other libraries.[53]

The efforts of the IFLA task force and many researchers have been focused on users' needs for powerful but easy-to-use search tools. While this goal is noble, there was nothing to guarantee that system vendors would comply with any of the suggestions and guidelines. As of yet, neither the three decades of OPAC display research nor the 1999 IFLA guidelines appear to have made much impact on the system design of integrated library systems. The only force that seems to hold any sway with vendors is the pursuit of greater market share.

Basic Search Queries

When approaching an unfamiliar information system, users must determine in which ways they can search. Author, title, subject, and keyword searches are found in most current retrieval tools. Some but not all OPACs also allow call number searches. In a few systems, the user may be offered searches that are less common, such as combined name/title, standard number, or journal title searches. While some search types may be fairly typical among systems, the ways in which systems are indexed are anything but consistent. To illustrate, consider some of the following questions that might arise as a user begins a search:

- Are all creators (primary and secondary; personal and corporate) included in the same creator index?
- How are creators searched?
- If keyword searching is available, which fields are searched?
- Can the records be searched for form or genre headings?
- How can users find works by *and* about a creator?

These are only a few of the questions a user might ask when beginning a new search. Due to the lack of standardization, the answers vary from system to system.

Users may or may not be able to choose whether a creator search is interpreted by the system as an exact match, browse, or keyword search. For efficiency, a user must know how the system searches a particular string. If a search for *Smith, William* is entered, will the system search for that word string exactly, no more and no less (i.e., a left-anchored phrase-matching search)? Will it search for any entry beginning with that exact word string (i.e., a browse search)? Will it search any creator-related field that contains both *William* and *Smith,* returning results for *Barney, William Smith* and *William, Jonathan Smith* (i.e., a keyword search)? In these examples, if a user wishes to see a record for a particular *William Smith* whose middle names and/or dates are not known or remembered, then a phrase match or a keyword match may not be satisfactory. Browsing may be the only efficient solution. Unfortunately, users may not understand the differences in these types of searches, and the system may not make these choices discernable or even available.

The results of keyword searching are determined by the fields that are searched in the system. Some systems have generic keyword searches that search almost every field of a record, or rather, a general keyword index is created by the system that includes everything but *stopwords* (a list of words that are so common that they are of no use when searching, such as *the*, *it*, *an*, *to*, and so on). In others, there may be separate buttons or pull-down menus so that users can choose the type of keyword search they want to perform (e.g., subject keyword, title keyword, author keyword). The fields to be included in these different types of keyword indexes vary from system to system. For example, some subject keyword indexes contain terms from subject heading fields, contents notes, and title fields, while others include only subject headings. Although note fields may contain subject-laden terminology, system designers almost never include them as fields to be searched for subject keywords.

Genre and form searching is relatively new and it is hoped that users will find this concept easy to understand. If users are interested in autobiographies or in specialized dictionaries, it is helpful for them to include the form or genre in the search. Few systems, so far, allow searching by genre/form, but as more metadata schemas incorporate the concept, the need for this type of search increases. For certain materials, like those used in the art and visual resources community, searches by genre/form are often essential to successful retrieval. In some systems, genre/form searching is not available, but genre/form browsing is becoming more available to users through the use of discovery interfaces (such as Primo, Summon, or VuFind) with various retrieval tools.

Over the years, it has become apparent that not all users understand the differences among searches by creator, title, or subject. This was discovered when card catalogs changed from the strict dictionary form to the divided catalog form. Users were especially confused by looking for persons *as authors* in the author drawers and looking for works *about* those same persons in the subject drawers. This problem was alleviated in some card catalogs by placing all works by and about persons in a *name* catalog. Very few online systems have allowed this solution, however. Users have to know that to find works about a person, one has to do a subject search, but to find works by a person, one has to do a creator search. Or they may search by keyword, but the results list is then a mixture of works by and about the person, along with anything else that might be retrieved from the database. Other searching considerations include choosing the collection or the library branch to search. Users should be aware also of the search limits available in a system (e.g., narrowing search results using language, date, format). As is true of most of the choices discussed here, all of these options are handled differently from system to system.

Initial Articles

Another example of the lack of consistency among systems is the handling of initial articles (*a, an, the,* and their equivalents in other languages). In many systems, there are instructions to omit initial articles when entering a search string. Users tend to follow this advice, *if* the instruction is noticeable and *if* it can be seen from the search box. If a search still includes the initial article, the system may

- return a message that the user received no hits, with no explanation given;
- provide the user with a message to remove the article and try again;
- treat an initial article differently depending on whether it is a keyword, browse, or exact word search; or
- eliminate the article without notifying the user and perform the search.

The last possibility in the above list may be good for most searches, but there are times when the article (or what appears to be an article) is needed for the search to be successful (e.g., a book titled *A is for Apple*).

Truncation, Boolean Operators, and Proximity

Other examples of the lack of standardization are found in the manner in which systems handle truncation, Boolean operators, and proximity. Users need to know if there is automatic right-hand truncation and, if not, which truncation symbol should be used. Automatic right-hand truncation happens when a user inputs a search word or string, and the system returns everything that begins with that word or string. For example, a search for *catalog* retrieves not only records for *catalog* but also records with the terms: *catalogue*, *catalogs, catalogues, cataloging, cataloguing, cataloged, catalogued, cataloger, cataloguer, catalogers, cataloguers,* and so forth. If right-hand truncation is not automatic, a symbol must be used to indicate that truncation is desired in a search. The symbol may be a pound sign (#), an asterisk (*), a dollar sign ($), or any other of a number of characters. For example, the search mentioned above may need to be formulated as *catalog#* in order to bring about the same results in a system without automatic truncation. The use of truncation might also change depending on the type of search conducted. For example, in a single system, title searches and creator searches may always use automatic truncation, but in keyword searches the user may need to input a specific truncation symbol.

Use of Boolean logic is far from consistent among systems. Most allow more than one word per command or search, but the ways in which multiple terms are treated can vary considerably. If a default Boolean operator is used, the default may be *AND* or it may be *OR.* Most online catalogs, for example, treat a search with multiple terms as if the operator *AND* has been inserted between them. That is, all terms in the search must be present in a record for that record to be retrieved. Many search engines, though, treat such a search as if *OR* has been inserted between terms. That is, each record retrieved contains at least one of the terms, but not necessarily all of the terms. When Boolean operators are *expressly* inserted, a user must know the order in which operations are carried out. For example, in a search for “catalogs OR indexes AND libraries,” most systems execute the operators from left to right (i.e., records with either one of the terms *catalogs* or *indexes,* combined with the word *libraries*). Some online retrieval systems, however, may search for records with the term *catalogs* OR records containing both *indexes* and *libraries.* The best solution would be to use nesting to ensure that the execution will be carried out as desired: “(catalogs OR indexes) AND libraries”—although this may be beyond the knowledge of many users. Adding to the confusion, Boolean operators may be allowed in some types of searches, but not in others.

Users also need to know if proximity formulations are supported and, if so, how specific the formulations are. Proximity formulations are often used in indexes and some search engines, but less often in library catalogs. The degree of sophistication in formulation may range from a simple *NEAR* or *ADJ* (adjacent) to an indication of the distance allowed to appear between terms (e.g., *w2* or *w/2*—meaning within two words).

Punctuation

Punctuation can be a source of great confusion for users. When approaching a new system, there are a number of questions that need to be asked. They include: Are quotation marks used to indicate exact phrases? Are diacritical marks used and understood? For example, are the accent marks in *résumé* necessary, and if so, how are they to be entered? Should all punctuation be stripped from the query, and if not, which symbols can be used and which ones must be deleted? Are there filing rules or conventions based on punctuation that will thwart a user’s ability to find desired search results? For example, are *meta-data* and *metadata* filed next to each other in a list of results or are they separated in the list because of the hyphen?

Removing punctuation from a string of text in order to provide better potential for matches is called *normalization.* Data normalization often occurs when the system indexes metadata records as they are entered into the database.

String:	**Benson, Susan Porter, 1943–2005.**
Normalized string:	**Benson Susan Porter 1943 2005**

Normalization is intended to allow better matching because users are not always precise in the placement of such marks as diacritics and commas. However, depending upon the algorithm used for normalization, results may be different from system to system.

A telling example of how the lack of standardization can affect users comes from an article by J. H. Bowman, who looked at how the ampersand (&) and the hyphen (-) are treated in various online catalogs.[54] Bowman found that the symbols were treated quite differently in discrete systems. The hyphen affected searching and the system’s indexing in at least six different ways. In some systems, it was treated like a space.

In others, it was treated as if it were a null character (i.e., nothing there, not even a space, thereby bringing the two pieces of the hyphenated word together into one word). Some systems treated it as both a null character and as a space (which meant the terms were indexed at least twice). Some viewed the hyphen simply as a hyphen, while one used the hyphen as a Boolean *NOT* (i.e., the hyphen as a minus sign). In one system, the hyphen invalidated the search altogether. In addition, Bowman found that the hyphen was sometimes processed differently in keyword searches than it was in phrase matching queries. The ampersand, too, had a variety of treatments. It was seen as an ampersand, as a Boolean *AND,* as a search string terminator, and as a null character. Besides the obvious effects that these different treatments have on the results and the arrangement of results lists (and on the frustration levels of users scrolling through long lists of results), they can also have an impact on federated searching.

Federated Searching and Z39.50

Federated searching, also known as *meta-searching,* is the ability to search and retrieve results from multiple sources of information while using only a single, common interface. It features an all-inclusive overlay (one-search box) to multiple systems, which may include catalogs, indexes, databases, and other electronic resources. William Frost states:

> Metasearching (a.k.a. federated searching or broadcast searching) is considered by some to be the next evolutionary step of database searching. Proponents believe that novice users, such as undergraduates, are baffled by the number of databases they have to choose from and need one common interface to meet all their research needs. A common interface, they claim, could wean novice searchers from using the Internet as their primary source for research.[55]

Despite the fact that many library websites default now to a single search box, federated searching is still a developing technology and the vocabulary has not, as of yet, been standardized. The meta-searching interface may be referred to as a common-user interface, an integrated interface, a standardized interface, or simply as a meta-database. Wendi Arant and Leila Payne give the following three conceptual models for meta-searching:

- a single interface that masks multiple databases and platforms
- a single, monster database with all sorts of data
- some combination of the two[56]

While it appears that users may desire this type of searching as a means of streamlining the research process, there are some downsides to consider. Federated searching functionality is limited to the level of the lowest common denominator. That is, if very sophisticated searching is available in System A but not in System B, then the searching that can be done from the more sophisticated System A is limited to the searching allowable in System B. This is true with the problems in truncation, Boolean operators, and punctuation discussed earlier. In addition, precision may be greatly reduced because authority control of names and subject terms may or may not be implemented in all of the information sources being searched, or the types of controlled vocabulary may differ completely. Frost states that, in addition to these reasons, reliance on meta-searching prevents students and other users from developing a better understanding of the multiple, diverse, specialized resources available to them and how to take advantage of the different systems.[57]

Steps toward resolving problems in federated searching were once centered on the use of the Z39.50 protocol when it seemed as though all library-related searching would be done through the OPAC. This idea is passé today. The OPAC is no longer most users' first stop when looking for information resources. Discussions of federated searching, now, are focused either on a meta-search engine (i.e., a search engine that searches multiple search engines) or on an interface superimposed on an integrated library system. In the latter, a single search box (found on the library's website) is the interface not only to the online catalog but also to numerous databases, electronic resources, and indexes made available by the institution. Although Z39.50 is not going to be at the center of federated searching in the future, it is not quite dead yet. It continues to play a role in how some integrated library systems interact with each other today.

The Z39.50 communication protocol is a national standard developed by the National Information Standards Organization (NISO). Z39.50 (the NISO number assigned to the first version of this standard) began as an attempt to get diverse library systems to communicate and share bibliographic information. The protocol establishes how one computer (the client) can query another computer (the server) and transfer search results from one to the other. In simple terms, using the Z39.50 protocol, one system translates a set of local commands into a set of universal commands that are then sent to another computer. The second system translates the universal commands into equivalent commands for the second system. The results are then returned to the first computer in the same fashion. If two institutions both have Z39.50 clients installed, OPAC users in one institution can search the catalog of the other institution using the commands of the local system with the display of results looking like the display used in the local system. Like federated searching, Z39.50 limits the searching of various catalogs to the lowest common denominator. The sophisticated programming of one system cannot be passed through Z39.50 to the other system.[58]

Although Z39.50 has been proven to be a useful protocol for communicating among information systems, developers are exploring new ways to perform the same tasks through updated processes. LC is now overseeing a new standard called *Search/Retrieval via URL* (SRU).[59] SRU is a standard XML search protocol for Internet search queries that uses the Contextual Query Language (CQL), a standard syntax for representing queries. It incorporates many of the ideas found in Z39.50 (i.e., it attempts to preserve many of the intellectual contributions that Z39.50 has accumulated in the last several decades); so, it can be viewed as an update or modernization of that protocol.

User-Centered System Design

Christine Borgman wrote that one of the reasons that online catalogs were hard to use was that the designers did not take into account users' needs. Anna Schulze states, "Most information professionals would agree that user-centered design makes an important contribution to high quality information systems. However, there is no general agreement about how to define the term *user-centered design* or how to best implement user-centered design strategies in the development of systems and services."[60] She goes on to say that user-centered design, depending on the situation, may refer to enhancing system performance to deliver better results; to creating a design for particular users, since one size does not fit all; or to understanding the user through continual user input into the design process. The three keys to user-centered design are (1) observation and analysis of users at work, (2) assistance from relevant aspects of design theory, and (3) iterative testing with users.[61] These are processes that have been incorporated into the field of information architecture, which is associated with designing and structuring web spaces. Andrew Dillon states, "conducting user-centered design does require specific skills and does involve methods and practices that shape designs in desirable ways. Information Architecture just happens to be a much better term for user-centered design and the creation of usable information spaces."[62]

Universal Design

Universal design (UD) is one approach to user-centered design. Universal design is a strategy that extends design to be as inclusive as possible. Sharon Farb states, "the goal of universal design is to accommodate the widest spectrum of users, which in turn increases commercial success."[63] Christopher Guder states, "UD is the practice of designing for all people, or as it is often defined, designing for barrier-free access, specifically with the intent of creating barrier-free access to . . . spaces for persons with a disability."[64] Without attention to universal design, information systems present multiple barriers to people with disabilities. For example, GUIs, multicolumn layout, and websites with frames may not be readable by users of speech and Braille devices; the computer mouse may not be operable for people with visual or certain orthopedic difficulties; audio cues cannot be heard by people with hearing loss; screen colors may obscure text for people with color blindness and for others with visual impairment; blinking cursors and flashing text may be detrimental to people with photosensitive epilepsy; and workstations may be inaccessible to people in wheelchairs. Farb advocates implementing universal design from the outset of a system design project to create products usable by all individuals, some of whom may encounter barriers in accessing information with the current level of technology implemented in many information systems.[65] Universal design enables *all* users to access the services and information provided.[66]

Multiple Languages/Scripts

User-centered design also is needed to assist people whose native languages and writing scripts vary from the ones used by the majority of users. Much material is available, particularly in research institutions, in multiple languages and scripts. Ever since library catalogers stopped handwriting catalog cards, there has been a problem producing records in such non-Roman languages as Japanese, Arabic, Chinese, Korean, Persian, Hebrew, and Yiddish (collectively known as JACKPHY languages), as well as Greek, Russian, and others. Romanization was developed as a way to write metadata in Roman characters and be able to interfile the records with those in English and other Roman-alphabet languages. However, a majority of native speakers of non-Roman languages have found it difficult or impossible to read the Romanized records or to search for what they need. As a result, extensive non-Roman language collections have been underutilized.

For more than two decades, records created in bibliographic networks have included vernacular characters or scripts (first for Chinese, Japanese, and Korean, and gradually adding others) along with the Romanized data, but systems have not been able to display the vernacular to the public. With the Unicode standard, some systems now have the ability to display records in many character sets.[67] Each character in every modern written language is represented with a 16-bit unit of encoding with no duplication. Users in countries where the Roman alphabet is predominant also need special software to be able to search using the other character sets in addition to seeing vernacular displays of records.

Other Aids for Users

Many recommendations for improving system design to benefit users have been made over the years. One of particular importance to users is that of spelling correction. It has been proposed that spell-check programs should be incorporated into the search process. Many systems now include normalization of search strings, which eliminates the need for users to input punctuation. However, research has shown that users often make spelling errors.[68] When users receive no results due to misspelled search words, they may walk away thinking the retrieval tool does not contain the information they were seeking. Some systems already assist with this. For example, Google asks a user who types in the word *excercise* if the user really meant to search *exercise* instead (although on the web, there are many hits for the misspelled word *excercise*). Other systems, such as those used in the airline industry, use sound-based codes so that names that sound alike but are spelled differently can be found easily. It has been suggested that such codes could be incorporated into information system design, along with dictionaries of alternative (e.g., American/British) spellings.

Other recommendations to improve system design for users have addressed some of the following issues:

- In many systems there are too few error messages, which are needed to help users understand their mistakes. Sometimes terminology used in error messages is not defined.
- Help screens are not always clear. And often the help system does not explain how to start or to refine a search.
- When users retrieve no hits or too many hits, instructions are needed to guide the user in how to increase or reduce results.
- Stopword lists differ from system to system. They differ in application (e.g., stopwords that apply only at the beginning of a search string versus stopwords that apply anywhere in the search string) and also differ in which terms are used as stopwords (e.g., *committee* is a stopword in one system, but not another).

Authority Control Integration

Authority control is a mechanism for creating consistency in online systems and for allowing greater precision and better recall in searching (and is addressed in detail in Chapter 8). *Precision* is a measurement of how many of the documents retrieved are relevant; *recall* is a measurement of how many of the relevant documents in a system are actually retrieved. Ideally, searchers would like both high precision and high recall. In other words, they would like to find all the resources that are exactly on their topic with very few non-relevant items in the mix. Precision is enhanced by the use of standardized forms of names, while recall is improved by the system of references created.

In order for authority control to be successful, authority work must be consistent and thorough. These days, in an integrated library system, authority records are linked to bibliographic records to ensure collocation of records that all relate to the same name, standardized title, or subject heading. These linkages may be one-to-one (e.g., an author who is associated with one information resource), or they may be one-to-many (e.g., a single author who has written several titles), or they may be many-to-one (e.g., linking more than one author to a single information resource). In subject headings, it is less common to find one-to-one relationships, as there are usually several or many works on the same topic. For a graphic representation of the relationship between bibliographic and authority records, see Figure 4.1. The figure shows the many-to-one and one-to-many relationships between authority records and bibliographic records.

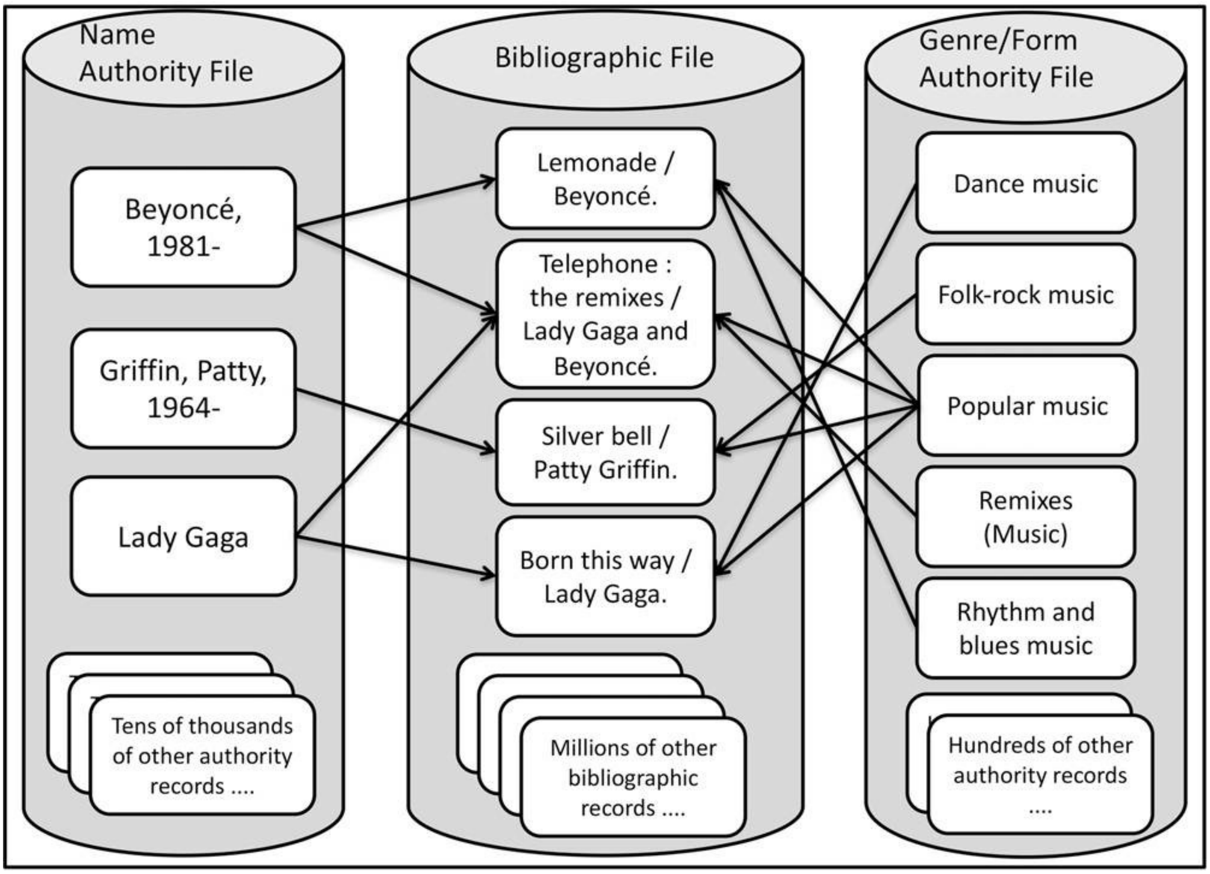


Figure 4.1. Linkages between Authority Records and Bibliographic Records.

The way this system linkage works, in simple terms, is this: Each authorized access point for a name (and for subjects and *some* titles) in a bibliographic record has an authority record. Each authority record may be linked to as many bibliographic records as are appropriate (e.g., the authority record for Lady Gaga is connected to the bibliographic record describing *Telephone*, to the one for *Born This Way*, and to every other record for her music), and each bibliographic record may be linked to as many authority records as are appropriate (e.g., the record describing *Telephone* is connected to two performers' authority records and two genre/form authority records). In the schematic representation in Figure 4.1, there are arrows from the authority records in the name authority file to the associated names in the records in the bibliographic file. This illustration is much simpler than what one would find in most catalogs, where a single bibliographic record may be linked to 10 or more different authority records for persons, corporate bodies, subject headings, genre/form headings, series titles, and so on.

In a system design there are links (sometimes called *pointers*) from each authority record to each bibliographic record that uses the name, title, or subject in that authority record (e.g., the bibliographic record number for *Lemonade* could be used as a pointer in the authority record for Beyoncé). There would also be links in each bibliographic record back to the multiple authority records that represent all the names, titles, and subjects for that record. The result is that if a user searches for an unused subject term, the authority record presents a reference to the authorized subject term used. Then when the user requests records for the used term (usually by clicking on a hyperlink in the display), all the bibliographic records using that term are displayed. Going in the other direction, if a user has found a bibliographic record that looks promising and wants more on that subject, the link from the subject back to the subject authority record allows the system to return all of the bibliographic records that use that subject.

Because authority work on names and controlled subject vocabularies is expensive, time-consuming, human-based work, keyword searching is often touted as being a sufficient method for retrieving information resources. However, research has shown that keyword searching may result in *false drops* (i.e., irrelevant retrievals) because the word retrieved has a different meaning from the intended meaning and also may result in loss of recall because synonyms or near-synonyms were not retrieved with the word sought. In addition, Tina Gross, Arlene G. Taylor, and Daniel N. Joudrey recently found that over one-quarter of keyword searches could fail if the authorized subject terminology is not included in the bibliographic records in the catalog, because the only place some keywords are found in some records is in the subject heading fields.[69] Users need to know, when searching, whether there are authority-controlled names, titles, and subjects, because, if not, users will have to think of all synonyms, related terms, and different forms of a name on their own.

Related to the need to know whether there are authority-controlled names and subject terms, users need to know whether subject relationships are available to help them browse a subject area, broaden or narrow searches, and look at related topics. In the area of subject searching there have been suggestions that keyword searches be matched against both free text terms and controlled vocabulary, and that system additions be made to assist users in taking advantage of the power of controlled vocabulary. An example comes from database indexes, which sometimes provide the ability to "explode a search." This allows a user to take advantage of the relationships among terms that are built into a controlled vocabulary. An *exploded search* is one in which the vocabulary term chosen is searched along with all of its narrower terms.

Another improvement to subject access might be to enhance subject terms with classification captions and terminology. Experimental systems have shown the viability of this approach. A notable one was the Cheshire II system created by Ray Larson and others.[70] In this system, words from the classification schedules that match the classification notation on the surrogate record are brought together with subject headings, subject-related title words, and subject-related terms from other parts of a surrogate record to create "subject clusters." Searching a cluster allows more accurate system responses than does searching each heading one by one. Others have suggested that user tags could work to supplement the subject headings as well. At this time, it seems rather unlikely that user tagging will enhance subject retrieval in any consistent and useful way, because the "problems of homographs, synonyms and polysemy, are common in them. Therefore, skepticism and ambiguity still exist in the professional cataloguing community about the value of social tagging."[71] It is, however, an area where research will likely continue as more Web 2.0 applications are incorporated into traditional system designs.

CONCLUSION

This chapter has discussed systems and system design as they relate to information organization. In the days of paper indexes and card catalogs, organizers were also the designers of the systems. Perhaps a goal today would be for organizers and system designers to become so conversant in each other's languages and knowledge that each could step into the other's position with ease (or, more realistically, at least have a conversation about what is needed). Sophisticated design is wasted if the metadata that the system presents is inadequate, and sophisticated metadata is wasted if the system design is lacking.

Information professionals have provided guidelines and suggestions for interface design and for the display of results and records. Researchers and practitioners have long aimed at resolving some of the problems that users encounter in retrieval systems. There are large bodies of research on information-seeking behavior, user needs, usability, the importance of subject access, and various retrieval methods. Some of the problems we see today could be eliminated if vendors would accept the guidelines and suggestions created by information organizations, professionals, and researchers. After 40 years of automated systems in organizing information, it is unfortunate that we must continue to ask the question, "Why are systems still hard to use?" There are ways to institute standard features in all systems that will not detract from the ability of vendors to create innovative and competitive products for the marketplace but will allow users to develop a lasting familiarity with retrieval tools. In order for users to get the most benefit from the information tools available, organizers and system designers must work closely together. This seems to be happening in the development and use of information discovery interfaces. These tools are allowing users to discover, browse, and explore resources in our collections. They show great promise in making accessible some resources that may have been hidden in traditional displays of retrieval results. Although there has been much progress in system design and in communication between organizers and designers, there is still a long way to go.

Some Important Terms in This Chapter

Auto-Graphics	Information Retrieval System	Phrase Matching
Bibliographic Network	Information System	Precision
Boolean Operator	Integrated Library System	Querying
Browsing	Interlibrary Loan	Recall
Copy Cataloging	Keyword Matching	Reference Database
Database	Knowledgebase	Retrospective Conversion
Discovery Interface	Known-Item Searching	SkyRiver
Facet	Left-Anchored Searching	Source Database
Faceted Browsing	Library Services Platform	Stopword
False Drop	OCLC Online Computer Library Center	System
Graphical User Interface	Online Public Access Catalog	System Design
Hierarchical Database	Original Cataloging	Turnkey System
Holdings Information		Universal Design

Some Important Acronyms in This Chapter

CQL: Contextual Query Language
GUI: Graphical User Interface
IFLA: International Federation of Library Associations and Institutions
ILL: Interlibrary Loan
ILS: Integrated Library System
ILMS: Integrated Library Management System
LC: Library of Congress
LHS: Library Housekeeping System
LMSS: Library Management Support System
LSP: Library Services Platform
MARC: MAchine-Readable Cataloging
NISO: National Information Standards Organization
NoSQL: Not Only Structured Query Language
OCLC: Online Computer Library Center
OPAC: Online Public Access Catalog
OSS: Open Source Software
SQL: Structured Query Language
SRU: Search/Retrieval via URL
UD: Universal Design
XML: Extensible Markup Language

NOTES

All URLs accessed June 2017.

1. *A Dictionary of Computing*, 4th ed. (Oxford: Oxford University Press, 1996), 489–90.

2. Larry Milsap, “A History of the Online Catalog in North America,” in *Technical Services Management, 1965–1990: A Quarter Century of Change, A Look to the Future: A Festschrift for Kathryn Luther Henderson*, eds. Linda C. Smith and Ruth C. Carter (Binghamton, NY: Haworth, 1996), 79–91.

3. “Database,” *FOLDOC: Free On-line Dictionary of Computing*, Denis Howe, ed., http://foldoc.org/database.

4. M. Jay Norton, “Knowledge Discovery in Databases,” *Library Trends* 48, no. 1 (Summer 1999): 9–21.

5. Neal Leavitt, “Will NoSQL Databases Live Up to Their Promise?” *Computer* 43, no. 2 (2010): 12–14.

6. OCLC, “OCLC Online Computer Library Center,” http://www.oclc.org/.

7. OCLC, “Inside WorldCat,” https://www.oclc.org/worldcat/inside-worldcat.en.html; OCLC, “About OCLC,” http://www.oclc.org/en/about.html.

8. Innovative Interfaces, “SkyRiver,” https://www.iii.com/products/metadata-management/#skyriver.

9. Auto-Graphics, “MARCit: Your Complete Cost Effective Cataloging Resource,” http://www4.auto-graphics.com/products-marcit-cataloging-resource.asp.

10. Carl Grant, “The Future of Library Systems: Library Services Platforms,” *Information Standards Quarterly* 24, no. 4 (2012): 5.

11. Marshall Breeding, “Serials Solutions to Launch Summon 2.0,” *Smart Libraries Newsletter* 33, no. 5 (May 2013): 2–5, https://journals.ala.org/index.php/sln/issue/view/281.

12. Marshall Breeding, “Library Services Platforms: A Maturing Genre of Products,” *Library Technology Reports* 51, no. 4 (2015): 8.

13. Carl Grant, “The Future of Library Systems,” 5.

14. Lucy A. Tedd, “OPACs through the Ages,” *Library Review* 43, no. 4 (1994): 27–37.

15. Milsap, “A History of the Online Catalog,” 79.

16. Ibid., 85–86.

17. Marshall Breeding, “Power Plays: Library Systems Report 2016,” *American Libraries* 47, no. 5 (May 2016): 36.

18. John Akeroyd and Andrew Cox, “Integrated Library Management Systems: Overview,” *Vine,* no. 115 (2000): 3–10.

19. Marshall Breeding, “Automation System Marketplace 2007: An Industry Redefined,” *Library Journal* 132, no. 6 (April 1, 2007): 36.

20. Oracle, “Oracle Buys Endeca,” https://www.oracle.com/corporate/acquisitions/endeca/index.html.

21. ProQuest, “AquaBrowser,” http://www.proquest.com/products-services/AquaBrowser.html.

22. Innovative Interfaces, “Encore,” https://www.iii.com/products/sierra-ils/encore-discovery/.

23. Ex Libris, “Primo,” http://www.exlibrisgroup.com/category/PrimoOverview; Ex Libris, “Summon Web-Scale Discovery,” http://www.exlibrisgroup.com/category/Summon_Web-Scale_Discovery.

24. LibLime, “Koha,” http://www.koha.org/.

25. Evergreen, “Evergreen: Open Source Library Software,” http://evergreen-ils.org/.

26. Villanova University Falvey Memorial Library, “VuFind: Search. Discover. Share,” http://vufind-org.github.io/vufind/.

27. Marshall Breeding, “Automation System Marketplace 2008: Opportunity Out of Turmoil,” *Library Journal* 133, no. 6 (April 1, 2008): 32.

28. Breeding, “Power Plays,” 37.

29. Current content from the magazine *American Libraries* may be found at https://americanlibrariesmagazine.org.

30. Charles R. Hildreth, “Online Catalog Design Models: Are We Moving in the Right Direction? A Report Submitted to Council on Library Resources August, 1995,” https://web.archive.org/web/20150923010236/http://myweb.cwpost.liu.edu/childret/clr-opac.html.

31. Akeroyd and Cox, “Integrated Library Management Systems,” 3.

32. Mary K. Bolin, “Catalog Design, Catalog Maintenance, Catalog Governance,” *Library Collections, Acquisitions, & Technical Services* 24 (2000): 53–63.

33. This example is from the derived searching used in OCLC, and is based on the prescribed formula for title searching of 3,2,2,1. That means the first three letters of the first non-stopword in the title, followed by the first two letters of the second word (whether stopword or not), and so on. An author/title derived search for the same book uses a 4,4 formula: joud,intr (i.e., the first four letters of the first author’s entry word [Joudrey] and the first four letters of the first non-stopword of the title [*Introduction*]).

34. Tedd, “OPACs through the Ages,” 28.

35. Hildreth, “Online Catalog Design Models.”

36. Ibid.

37. For example: Karen G. Schneider, “How OPACs Suck, Part 2: The Checklist of Shame,” in “ALA TechSource,” http://www.ala.org/tools/article/ala-techsource/how-opacs-suck-part-2-checklist-shame.

38. EBSCO, “EBSCO Discovery Service,” https://www.ebscohost.com/discovery.

39. Minuteman Library Network, “Catalog,” https://find.minlib.net/iii/encore/?lang=eng.

40. Library Thing, http://www.librarything.com/.

41. University of Illinois at Urbana-Champaign, “Library Catalog,” https://vufind.carli.illinois.edu/vf-uiu/ or http://www.library.illinois.edu/.

42. Robin Hastings, “Linked Data in Libraries: Status and Future Directions,” *Computers in Libraries* 35, no. 9 (2015): 15, http://www.infotoday.com/cilmag/nov15/Hastings-Linked-Data-in-Libraries.shtml.

43. Hildreth, “Online Catalog Design Models.”

44. Christine L. Borgman, “Why Are Online Catalogs Hard to Use? Lessons Learned from Information Retrieval Studies,” *Journal of the American Society for Information Science* 37, no. 6 (June 1986): 387–400.

45. Christine L. Borgman, “Why Are Online Catalogs *Still* Hard to Use?” *Journal of the American Society for Information Science* 47, no. 7 (July 1996): 493–503.

46. Ibid., 501.

47. David Thomas, “The Effect of Interface Design on Item Selection in an Online Catalog,” *Library Resources & Technical Services* 45, no. 1 (January 2001): 20–45.

48. This was reinforced by LC selling its cataloging on “3 × 5” cards. The size of the LC cards actually was, and continued to be, 7.5 cm × 12.5 cm, but the United States did not change to the metric system as Dewey believed was imminent. The size was just under 3 in. × 5 in., and the cards came to be called “3 × 5” cards.

49. Martha M. Yee, “Guidelines for OPAC Displays: 1999,” in *From Catalog to Gateway: Charting a Course for Future Access: Briefings from the ALCTS Catalog Form and Function Committee,* eds. Bill Sleeman and Pamela Bluh (Chicago: Association for Collections & Technical Services, American Library Association, 2005), 83, also available at http://www.pitt.edu/~agtaylor/ala/papers/YeeOPACGuidelines.pdf.

50. Allyson Carlyle and Traci Timmons, “Default Record Displays in Web-Based Catalogs,” *Library Quarterly* 72, no. 2 (April 2002): 179–204.

51. Thomas, “The Effect of Interface Design,” 42–44.

52. Ibid., 32.

53. Yee, “Guidelines for OPAC Displays,” 84.

54. J. H. Bowman, “The Catalog as Barrier to Retrieval—Part 1: Hyphens and Ampersands in Titles,” *Cataloging & Classification Quarterly* 29, no. 4 (2000): 39–60.

55. William F. Frost, “Back Talk: Do We Want or Need Metasearching?” *Library Journal* 129, no. 6 (April 1, 2006): 68.

56. Wendi Arant and Leila Payne, "The Common User Interface in Academic Libraries: Myth or Reality?" *Library Hi Tech* 19, no. 1 (2001): 63–76.

57. Frost, "Back Talk: Do We Want or Need Metasearching?" 68.

58. Paul Miller, "Z39.50 for All," *Ariadne* 21 (September 20, 1999), http://www.ariadne.ac.uk/issue21/z3950/.

59. Library of Congress, "SRU: Search/Retrieval via URL: SRU (the Protocol), CQL (Query Language), ZeeRex (Service Description)," http://www.loc.gov/standards/sru/.

60. Anna Noakes Schulze, "User-Centered Design for Information Professionals," *Journal of Education for Library and Information Science* 42, no. 2 (Spring 2001): 116.

61. Ibid.

62. Andrew Dillon, "Information Architecture in JASIST: Just Where Did We Come From?" *Journal of the American Society for Information Science and Technology* 53, no. 10 (2002): 822, https://www.ischool.utexas.edu/~adillon/Journals/IA%20ASIST%20intro.pdf.

63. Sharon Farb, "Universal Design and the Americans with Disabilities Act: Not All Systems Are Created Equal—How Systems Design Can Expand Information Access: 2000," in *From Catalog to Gateway: Charting a Course for Future Access: Briefings from the ALCTS Catalog Form and Function Committee,* eds. Bill Sleeman and Pamela Bluh (Chicago: Association for Collections & Technical Services, American Library Association, 2005), 99–106, also available at http://www.pitt.edu/~agtaylor/ala/papers/FarbUniversalDesign.pdf.

64. Christopher Guder, "Universal Design and the Library Website," *Public Services Quarterly* 10, no. 3 (2014): 229.

65. Sharon Farb, "Universal Design and the Americans with Disabilities Act."

66. Guder, "Universal Design and the Library Website," 236.

67. "The Unicode Consortium," http://www.unicode.org/.

68. Arlene G. Taylor, "Authority Files in Online Catalogs: An Investigation of Their Value," *Cataloging & Classification Quarterly* 4, no. 3 (1984): 1–17.

69. Tina Gross, Arlene G. Taylor, and Daniel N. Joudrey, "Still a Lot to Lose: The Role of Controlled Vocabulary in Keyword Searching," *Cataloging & Classification Quarterly* 53, no. 1 (2015): 1–39.

70. Ray R. Larson, Jerome McDonough, Paul O'Leary, and Lucy Kuntz, "Cheshire II: Designing a Next-Generation Online Catalog," *Journal of the American Society for Information Science* 47, no. 7 (July 1996): 555–67.

71. Praveenkumar Vaidya and N. S. Harinarayana, "The Comparative and Analytical Study of LibraryThing Tags with Library of Congress Subject Headings," *Knowledge Organization* 43, no. 1 (2016): 41.

SUGGESTED READINGS

Bates, Marcia J. "The Design of Browsing and Berrypicking Techniques for the Online Search Interface." *Online Review* 13, no. 5 (October 1989): 407–24.

Breeding, Marshall. "Library Services Platforms: A Maturing Genre of Products." *Library Technology Reports* 51, no. 4 (2015): 1–40.

Breeding, Marshall. "Library Systems Report." *American Libraries.* May 1st issue each year.

Breeding, Marshall. *Library Technology Guides*. http://www.librarytechnology.org/.

Breeding, Marshall. *Next-gen Library Catalogs*. New York: Neal-Schuman, 2010.

Chambers, Sally, ed. *Catalogue 2.0: The Future of the Library Catalogue*. Chicago: Neal-Schuman, 2013.

Enis, Matt. “All Systems Go: Library Systems Landscape 2016.” *Library Journal* 140, no. 6 (2016): 34–38.

Enis, Matt. “Library Systems Landscape 2016: Company Profiles.” *Library Journal* 140, no. 6 (2016): 42–44.

Farb, Sharon. “Universal Design and the Americans with Disabilities Act: Not All Systems Are Created Equal—How Systems Design Can Expand Information Access: 2000.” In *From Catalog to Gateway: Charting a Course for Future Access: Briefings from the ALCTS Catalog Form and Function Committee,* edited by Bill Sleeman and Pamela Bluh, 99–106. Chicago: Association for Collections & Technical Services, American Library Association, 2005. Also available at http://www.pitt.edu/~agtaylor/ala/papers/FarbUniversalDesign.pdf.

Gallagher, Matt. “How to Conduct a Library Services Platform Review and Selection.” *Computers in Libraries* 36, no. 8 (2016): 20–22.

Gross, Tina, Arlene G. Taylor, and Daniel N. Joudrey. “Still a Lot to Lose: The Role of Controlled Vocabulary in Keyword Searching.” *Cataloging & Classification Quarterly* 53, no. 1 (2015): 1–39.

Hernandez, Michael J. *Database Design for Mere Mortals: A Hands-on Guide to Relational Database Design*. 3rd ed. Upper Saddle River, NJ: Addison-Wesley, 2013.

Hildreth, Charles R. “The Use and Understanding of Keyword Searching in a University Online Catalog.” *Information Technology and Libraries* 16, no. 2 (June 1997): 52–62.

Joudrey, Daniel N., Arlene G. Taylor, and David P. Miller. *Introduction to Cataloging and Classification*. 11th ed. Santa Barbara, CA: Libraries Unlimited, 2015. Chapter 24: “Cataloging Management and Support.”

Larson, Ray R. “Classification Clustering, Probabilistic Information Retrieval, and the Online Catalog.” *Library Quarterly* 6, no. 2 (April 1991): 133–73.

Larson, Ray R., Jerome McDonough, Paul O’Leary, and Lucy Kuntz. “Cheshire II: Designing a Next-Generation Online Catalog.” *Journal of the American Society for Information Science* 47, no. 7 (July 1996): 555–67.

Morville, Peter, and Jeffery Callender. *Search Patterns*. Sebastopol, CA: O’Reilly Media, 2010.

OCLC. *Online Catalogs: What Users and Librarians Want: An OCLC Report*. Dublin, OH: OCLC, 2009.

Pace, Andrew. “21st Century Library Systems.” *Journal of Library Administration* 49, no. 6 (2009): 641–50.

Schulze, Anna Noakes. “User-Centered Design for Information Professionals.” *Journal of Education for Library and Information Science* 42, no. 2 (2001): 116–22.

Wang, Yongming, and Trevor A. Dawes. “The Next Generation Integrated Library System: A Promise Fulfilled.” *Information Technology and Libraries* 31, no. 3 (2012): 76–84.

Webber, Desiree, and Andrew Peters. *Integrated Library Systems: Planning, Selecting, and Implementing*. Santa Barbara, CA: Libraries Unlimited, 2010.

CHAPTER 5

INTRODUCTION TO METADATA

In the preceding chapters, we have referred briefly to the concept of metadata in numerous contexts. In this chapter, we go into greater detail about metadata and its types, forms, characteristics, and uses. We look at some useful metadata management tools, some conceptual models that shape the metadata creation process, and how metadata helps to meet the objectives of information retrieval tools.

Metadata is commonly described as "data about data." This definition assumes that an information resource (a book, a map, a blog, a photo on Instagram, a streaming video, etc.) is a form of data. A description of the attributes and contents of that resource would *also* be data; hence the definition "data about data." This definition represents the very broadest level of the concept. In a cursory survey of terminology usage, numerous definitions can be found, ranging from the aforementioned "data about data" to more complex, lengthier definitions. What they all have in common is the notion that metadata is structured information that describes the important attributes of information resources for the purposes of identification, discovery, selection, use, access, and management. *FOLDOC: Free On-Line Dictionary of Computing* defines metadata as "definitional data that provides information about or documentation of other data managed within an application or environment. . . . Metadata may include descriptive information about the context, quality and condition, or characteristics of the data."[1] This definition implies that metadata includes not only descriptive information, such as that found in traditional retrieval tools for the purpose of resource discovery, but also information necessary for the management, use, and preservation of the information resource (e.g., data about where the resource is located, how it is displayed online, its ownership, its condition).

Even among information professionals, metadata concepts can be complex and confusing. This is due in part to the multifaceted nature of the topic and in part to the overly broad, pervasive data-about-data definition. With that as a primary description, it is no wonder that many refer to any number of interrelated concepts as *metadata*. Discussions about metadata may be about any one (or any combination) of the conceptual components listed in Table 5.1.

While in theory we may discuss any or all of these as distinct conceptual components, in practice the divisions among them are not quite so clear. At times, the concepts are so intertwined that efforts to separate them only result in more confusion. An example to illustrate this can be found in the MAchine-Readable Cataloging (MARC) bibliographic format. While many consider MARC to be nothing more than an encoding format, others refer to it as a metadata schema (or structure standard). The confusion stems from MARC's exhibiting properties of both encoding formats and metadata schemas. On closer examination, MARC even acts as a data value standard by dictating the values of certain data elements in the record (particularly the MARC control fields). So, it is not surprising that there are misunderstandings about what metadata is. The components addressed in Table 5.1 are addressed in later chapters of this book.[2]

It is important to remember that the term *metadata* may mean different things to different communities. The information resources found in libraries are quite different from those found in museums, historical societies, repositories of scientific data sets, or on the web. Differences in the nature, characteristics, and uses of the resources may require diverse approaches to description and to the metadata created. Different types of information professionals (and nonprofessionals of all types) will therefore have different notions of what metadata entails. For example, when librarians, who are mostly familiar with the bibliographic data found in a catalog, speak of metadata, they often have a different notion of the concept than someone who designs databases or works with complex geospatial information encoded in an Extensible Markup Language (XML) schema.

THE BASICS OF METADATA

Metadata may be divided into three broad categories: *administrative metadata, descriptive metadata,* and *structural metadata.*[3] These categories, however, are somewhat fluid. It is important to remember that the boundaries among them are not fixed and their definitions are not standardized or precise. Individual metadata elements may fit into more than one of these categories (e.g., a unique identifier may be considered useful as descriptive, administrative, and structural metadata). In addition, these three categories are not necessarily the only ways to classify metadata. Some authors have identified five types of metadata.[4]

Table 5.1. Components That May Be Referred to as Metadata.

Component	Explanation	Examples
Attributes of Information Resources	**The identifying characteristics of resources (i.e., the data used to describe them)**	• ***Jaws*—the title of a film** • ***2015*—the manufacture date of a specimen collection** • ***Bungee jumping*—the subject of a resource**
Elements	**Individual units of information that contain the specific attributes used to describe resources**	• **<title>** • **<date>** • **<subject>**
Data Structure Standards *(or Metadata Schemas)*	**Sets of metadata elements created by particular communities or for particular types of resources**	• **Dublin Core Metadata Element Set (DCMES)** • **International Standard Bibliographic Description (ISBD)** • **Visual Resources Association's (VRA) Core Categories**
Data Content Standards	**Sets of rules, guidelines, or best practices used to create resource descriptions**	• ***RDA: Resource Description & Access*** • ***Describing Archives: A Content Standard* (DACS)** • ***Cataloging Cultural Objects* (CCO)**
Data Value Standards *(or Controlled Vocabularies)*	**Lists of controlled values usedto populate particular metadata elements**	• ***Library of Congress Subject Headings* (LCSH)** • **Getty's *Union List of Artist Names* (ULAN)** • **Dublin Core Metadata Initiative (DCMI) Type Vocabulary**
Data Format Standards *(or Encoding Standards)*	**Syntaxes used to convert resource descriptions into machine-readable form**	• **MAchine-Readable Cataloging (MARC) bibliographic format** • **Encoded Archival Description (EAD)**
Models	**Holistic representations of the principles, entities, attributes, and relationships used to describe information resources**	• ***Functional Requirements for Bibliographic Records* (FRBR)** • ***Conceptual Reference Model* of the International Council on Museums Committee on Documentation (CIDOC-CRM)** • ***IFLA Library Reference Model* (LRM)**
Records	**Documented descriptions of (or surrogates for) information resources**	• **A catalog record** • **A finding aid** • **A metadata statement**

- Administrative
- Descriptive
- Preservation
- Technical
- Use

Others, though, have identified as many as 11 metadata types.[5]

- Administrative
- Behavior
- Descriptive
- Image quality assessment
- Meta-metadata
- Preservation
- Records management or Recordkeeping

- Rights
- Structural
- Technical
- Tracking

Some authors have offered additional types of metadata to consider, such as *contextual metadata*[6] and *analytical metadata*.[7] Others, however, view these additional types as part of the descriptive metadata category. We must remember that because there is no formal taxonomy of metadata types with precise definitions, what one author might refer to as *technical metadata* another might call *structural metadata* or *administrative metadata.* The three broad categories employed in this text reflect the most prevalent categorization used today. Some of the more frequently encountered types listed above (technical, rights, preservation, and meta-metadata) are included as sub-types of administrative metadata. Each is addressed later in this chapter. The others are subsumed under the three main categories. For a broad overview of these metadata types, see Table 5.2.

Table 5.2. Metadata Types.

Type	Definition	Sub-types	Typical Elements	Examples
Administrative Metadata	Supports management, decision making, preservation, rights management, technical support, metadata management, usage tracking, assessment, and record keeping	• Image-quality assessment • Meta-metadata • Preservation • Recordkeeping • Rights • Technical • Tracking • Use	• Acquisition • Condition • Digitization information • File size • Metadata creation data • Preservation actions • Responsibility • Software requirements • Storage requirements • Usage information • User tracking	• Metadata for Images in XML Standard (MIX) • Open Digit al Rights Language (ODRL) • Preservation Metadata: Implementation Strategies (PREMIS)
Structural Metadata	Technical information that is needed to ensure that a digital resource functions properly, displays onscreen, and can be navigated by users	• Behavior	• Behavior • File sequencing • Next page • Previous page • Resource map	• Material eXchange Format (MXF) • Metadata Encoding & Transmission Standard (METS) • Open Archives Initiative Object Reuse and Exchange (OAI-ORE) • Page-turner models
Descriptive Metadata	Describes the identifying characteristics and intellectual contents of information resources for the purposes of discovery, identification, selection, acquisition, context, and understanding	• Analytical • Contextual	• Contributors • Creator or Author • Dates • Edition or Version • Notes • Provenance • Series • Standard numbers • Style • Subject • Table of contents • Title	• Categories for the Description of Works of Art (CDWA) • Dublin Core Metadata Element Set (DCMES) • Friend of a Friend (FOAF) • RDA elements • Schema.org • VRA Core Categories

Another way of looking at metadata is to consider its level of complexity. Metadata can be classified into three levels. The first is a simple format, in which the metadata is really no more than data extracted from the resource itself. This is reflected by the search engine approach to organizing the web through automated indexing techniques. The second level is a structured format. This includes formal metadata element sets that have been created for the general user. This level of metadata may have a basic template for metadata creation and does not require professional-level description. The use of Hypertext Markup Language (HTML) meta tags reflects this level of complexity. The third level of metadata is a rich format. Information professionals in libraries, archives, and museums tend to use this third level to create comprehensive, meticulously detailed descriptions. They are more complex in nature and combine metadata elements with encoding and content standards. Examples of rich formats are found in bibliographic records that are created using *RDA: Resource Description & Access*, International Standard Bibliographic Description (ISBD), and MARC, and in online finding aids created using *Describing Archives: A Content Standard* (DACS) and Encoded Archival Description (EAD). At times, the need for such comprehensive, rich description has been questioned by those interested in developing more cost-effective approaches to metadata creation, often advocating for greater use of automatically generated metadata whenever possible. Others, however, are convinced that this will result in a reduction of the quality of the metadata produced and are opposed to making radical changes in metadata-creation processes. As the popular saying goes (popular among catalogers and archivists, that is), "Metadata is a love-letter to the future." In other words, many strongly believe it is better to create rich, full, well-formed metadata now, than to regret the lack of it later—we should be kind to future generations.

Metadata can be used to describe information resources at various levels of *granularity* (i.e., the level of detail or depth). It can be created for individual information resources, for discrete components of those resources, or for established collections comprising multiple resources. In other words, metadata could be used to describe

- a single web page, which may contain text, images, one or more multimedia files, or any combination of these;
- an individual streaming video clip embedded on that page or one of the other components; or
- the entire website that contains the page with the video clip, as well any number of other pages that are part of the site.

Different communities may choose to describe their resources at different levels of granularity. Some communities might describe resources at any or all of these levels, depending on the type of resource, the community's approach to organizing information, and the needs of their users. For example, libraries typically describe printed books on the individual resource level (i.e., one book equals one metadata record). Archival materials, however, are traditionally described with less granularity, usually at the collection level. Digital libraries may combine or alternate between collection-level and item-level descriptions, based on the nature of a particular collection, the size of the collection, and the users of the collection. In other words, a digital library might describe individual objects (e.g., a digitized map), individual collections (e.g., a collection of 2,000 digitized maps), or both.

Metadata may be found in various places and in different forms. For example, the metadata for an online resource may be part of the document itself (entered in the header as meta tags) or it may be found as embedded annotations within the HTML coding (using microdata to mark up web documents). Metadata found in these forms may be viewed by looking at the encoding for the document, and it may or may not be indexed by an Internet search engine. Metadata may also be found "wrapped" into a digital object's complex packaging. The descriptive metadata may be one small section of a multi-file digital object (such as a digitized music album comprising 10 tracks, cover photos, liner notes, searchable lyrics, etc.), and may only be searched and viewed within the content management system used to house the digital repository. Metadata may also be viewed as a surrogate record that is separate from the information resource that it describes. These records may be held in various retrieval tools to allow users to browse or search for the records, instead of trying to navigate through each individual resource in the collection. For online materials, there may or may not be direct links to the individual resources from their surrogate records, depending on the nature of the resource, the institution, and the sophistication of the retrieval tool. Because metadata is usually created in the electronic environment (e.g., in an online catalog, in an XML database, or on the web), the term is rarely applied to records found in paper tools.

METADATA SCHEMAS

In order for it to be used to its full potential, metadata cannot consist of unstructured descriptions of resources; it must be standardized and controlled. Without formal structure or rules, metadata description is no better than keyword access. Two basic units of metadata are the element and the schema. *Metadata elements* are the individual categories or fields that hold the discrete pieces of a resource description. Typical metadata elements include title, creator, unique identifier, creation date, subject, and the like. *Metadata schemas* are sets of elements designed to meet the needs of particular communities. While some schemas are general in nature, most are created for specific types of information resources or for specific purposes. Specific schemas have been designed to manage government, geospatial, visual, educational, and other types of resources. Consequently, it is not surprising that these schemas vary greatly. They differ in the number of data elements, in the use of mandatory and repeatable elements, in encoding requirements, and in the use of value standards (i.e., controlled vocabularies), among other things. While most schemas focus on descriptive elements to support resource discovery and identification, some contain elements to support administrative and structural purposes. Much to the chagrin of students and information professionals alike, it is simply not possible to create a perfect, one-size-fits-all metadata schema that will suffice for the various needs of different communities. Diversity in metadata is both desirable and unavoidable. The goal is to find ways for these different communities to share their useful, meaningful metadata.

According to Sherry Vellucci, there are three characteristics found in all metadata schemas. They are semantics, syntax, and structure.[8] *Semantics* refers to meaning, specifically the meaning of the various metadata elements. Semantics help metadata creators understand, for example, what the element *coverage* means in a given schema. *Syntax* refers to the encoding of the metadata. This may be the MARC format for bibliographic records or an XML schema for other types of metadata (e.g., EAD for an archives' finding aids). Stuart Weibel writes that syntax allows us to "take a set of metadata assertions and pack them so that one machine can send them to another, where they can be unpacked and parsed. . . . Syntax is arranging the bits reliably so they travel comfortably between computers."[9] *Structure* refers to the data model used to shape the way that metadata statements are expressed. Weibel states, "structure is the specification of the details necessary to layout [*sic*] and declare metadata assertions so they can be embedded unambiguously in a syntax. A data model is the specification of this structure."[10] The structure of metadata should not be confused with *structural metadata* (described elsewhere in this chapter), which refers to the organization of the resource being described. *Structure*, on the other hand, is referring to the expression of metadata statements (e.g., whether the metadata conforms to the Resource Description Framework [RDF] structure described below).

The semantics of a metadata schema do not necessarily dictate the content placed into the elements. This is the province of content standards and data value standards. *Content standards* are used to dictate things such as how dates will be formatted within metadata elements or how a personal name will be entered. For example, a set of best practices guidelines might state that all dates are to be recorded using the YYYY-MM-DD format; or a content standard may require personal names to be entered with family name first, followed by a comma, and then by the remainder of the name. Controlled vocabularies also may be used to standardize the contents of elements such as subject or resource type. *Controlled vocabularies* are lists of words in which certain terms are chosen as preferred, and their synonyms act as pointers to the preferred terms, thereby limiting the range of values that can be entered into a particular metadata element. If tools such as content standards and data value standards did not exist, information retrieval would be less consistent and less effective.

METADATA CHARACTERISTICS

In order for metadata to be as useful as possible in meeting the diverse needs of information users, some specific metadata characteristics require attention: interoperability, flexibility, and extensibility.[11] *Interoperability* refers to the ability of various systems to interact with each other no matter the hardware or software being used. Achieving interoperability helps to minimize the loss of information due to technological differences. Interoperability can be divided into semantic, syntactic, and structural interoperability. *Semantic interoperability* refers to ways in which diverse metadata schemas express meaning in their elements. In other words, does the element *author* in one schema mean exactly or nearly the same thing as *creator* in another? *Syntactic interoperability* refers to the ability to exchange and use metadata from other systems. Syntactic interoperability requires a common encoding format (e.g., Can a MARC record be used and understood in an XML environment?). *Structural interoperability* refers to how metadata statements are expressed. Is the metadata understandable by other systems? Are both systems using an entity-relationship model or are the basic structures different? Without interoperability on all three levels, metadata cannot be shared effortlessly, efficiently, or profitably.

Flexibility refers to the ability of "metadata creators to include as much or as little detail as desired in the metadata record, with or without adherence to any specific cataloging rules or authoritative lists."[12] Some information professionals seek to create flexible metadata schemas that leave decisions about how much detail to include in the description to the individual institution or community. Flexibility, at least at some level, is seen in most metadata schemas, especially those that contain many optional elements that are included at the discretion of the agency or the individual creating the metadata content.

Extensibility refers to the ability to use additional metadata elements and *qualifiers* (i.e., element refinements) to meet the specific needs of various communities. Qualifiers help to refine or sharpen the focus of an element or might identify a specific controlled vocabulary to be used in that element (e.g., the element *title* could have a qualifier to specify *subtitle*, *alternative title*, *earlier title*, or another variant form). An example of extensibility can be found in the education community. In order to meet their specific needs, a standard element set was extended by adding *instructional method* and *audience* as elements. There is a note of caution about extensibility, though. As extensibility increases, interoperability can decrease, because as the schema moves further away from its original design (through additional elements or qualifiers), it becomes less understandable to other systems using the base schema. In other words, there is an inverse relationship between extensibility and interoperability. This trade-off must be considered carefully before being implemented.

CATEGORIES OF METADATA

As mentioned above, in this text we are using three broad categories of metadata: *descriptive*, *administrative*, and *structural*. Until recently, most discussions of metadata in the library and information science (LIS) professions focused solely on descriptive metadata. This is unsurprising given that description is (and has been) a familiar activity; it is something that catalogers, indexers, technical services archivists, and museum registrars do regularly. Although administrative metadata is not new—recordkeeping has been around for ages—traditionally, it has not received as much attention as descriptive metadata. It is only within recent years that structural metadata needs have been recognized, or at least acknowledged, as a metadata issue. Each of these three broad categories of metadata is described below.

Descriptive Metadata

Descriptive metadata contains the most important identifying characteristics of an information resource and the analysis of its contents for the purposes of discovery, identification, selection, and acquisition. It includes the following kinds of information:

- **Identifying data**: titles, statements of responsibility, dates of publication or distribution, languages, formats, resource types, identifiers;
- **Relationship data**: creators, contributors, publishers, sources, related works, identification of other relationships among the various entities; and
- **Content data**: subjects, abstracts, coverage, tables of contents, classification notations, and categories.

This section is brief because Chapters 7–11 cover various aspects of descriptive metadata in much more depth.

Administrative Metadata

If an institution creates a digital resource, where is it held? How is it stored? What type of object is it? Who decides what it is called and when it needs to be updated? How are these processes accomplished? How did the object come into the collection? Was it born digital? Was it digitized in-house, by a vendor, or by another organization? Who can access it and for what purposes? These questions can be answered with administrative metadata.

Administrative metadata is created for the purposes of management, decision making, and record keeping. It provides information about the technical, preservation, and storage requirements of information resources, particularly digital objects. Administrative metadata assists with the monitoring, accessing, reproducing, digitizing, and backing up of digital resources. It includes information such as

- **Hardware and software requirements**: creation software, creation hardware, operating system;
- **Acquisition information**: when the resource was created, modified, and/or acquired; administrative information about the analog source from which a digital object was derived;

- **Ownership, rights, permissions, legal access, and reproduction information**: what rights the organization has to use the material, who may use the material and for what purposes, what reproductions exist and their current status;
- **Use information**: what materials are used, when, in what form, and by whom; use tracking and circulation statistics; user tracking; content re-use; exhibition records;
- **File characteristics**: file size, bit-length, format, presentation rules, sequencing information, running time, file compression information;
- **Version control**: what versions exist and the status of the resource being described; alternate digital formats, such as HTML or PDF for text, and GIF or JPG for images;
- **Digitization information**: compression ratios, scaling ratios, date of scan, resolution;
- **Authentication and security data**: inhibitor, encryption, and password information; and
- **Preservation information**: integrity information, physical condition, preservation actions, refreshing data, migrating data, conservation or repair of physical artifacts.

Some administrative metadata elements, particularly technical metadata such as file sizes and compression rates, may be generated automatically when the object is created or digitized. A familiar example is the Exchangeable Image File Format (Exif) metadata automatically created by a camera when a digital photograph is taken. Other elements, however, need to be recorded manually.

Unlike most descriptive metadata, administrative metadata is not standardized. As of this time, there is no single administrative metadata schema that has been adopted widely among institutions. The metadata captured for management purposes, such as who makes decisions and their contact information, tends to be repository specific and is stored in a variety of forms, in a variety of places (e.g., administrative metadata may or may not be incorporated into a record primarily comprising descriptive metadata). While there is no ubiquitous general administrative metadata schema, some separate schemas have been created to address specific sub-types of administrative metadata, such as Preservation Metadata: Implementation Strategies (PREMIS)[13] and NISO Metadata for Images in XML Schema (MIX)[14] for technical metadata for digital images.

As stated above, administrative metadata may be described as having several sub-types. These include: technical metadata, preservation metadata, rights and access metadata, and meta-metadata. Each is discussed below.

Technical Metadata

If an information institution receives a digital resource, would they know what it is? Would they understand what it can do? Could they interact with it or make it work? Without technical metadata, they might not be able to make use of the object.

According to PREMIS, "Technical metadata describes the physical rather than intellectual characteristics of digital objects."[15] Basic technical information is needed in order to understand the nature of the resource, the software and hardware environments in which it was created, and what is needed to make the resource accessible to users. Technical metadata describes the characteristics, origins, and lifecycles of digital documents, and is key to the preservation of the resource for future use. Because technical metadata is format specific, different schemas are used for different resource types (e.g., a streaming video requires different technical metadata than a digital image because they work in different ways and have different characteristics). As mentioned above, a typical example of technical metadata is that which is automatically attached to a digital photograph when it is taken (see Figure 5.1). While some of the metadata elements may overlap, technical metadata is *not* the same thing as structural metadata. The biggest difference between the two is that structural metadata is primarily machine readable and machine processable, while technical information is for humans to read and understand.

Preservation Metadata

Technology changes rapidly. Data files from just a few years ago—our WordStar documents and VisiCalc spreadsheets—are now old and no longer functional. Many have been lost because there were no plans to keep the data usable and no real understanding of how quickly technology changes. With this in mind, we must consider how much more we are willing to lose. Will the MS Word documents and Excel spreadsheets we create today be usable in 200 years? What about 20 years? Or five years, even? To save this information from oblivion, we must consider a number of questions. Should we preserve the look and feel of the software, or are we concerned with content only? One hundred years from now, will users still know what a *.jpg* or a *.tiff* is? Will they be decipherable? When we preserve an information resource, how do we determine which version of that resource needs to be preserved? In other words, how do we determine the best version? These questions are preservation metadata concerns.

In the past two decades, the Library of Congress, OCLC, the Research Libraries Group (RLG), and the National Library of Australia, among others, launched preservation metadata initiatives. Preservation metadata is the information needed to ensure the long-term storage and usability of digital content. It may include information about reformatting, migration, emulation, conservation, file integrity, and provenance. Typical preservation metadata elements might include the following:

File and Information	
Name:	530541_10200170393760766_1093851742_n.jpg
Kind:	JPEG image
Size:	880 KB (879,746 bytes)
Created:	Nov 03, 2008, 04:11:59 pm
Modified:	Nov 03, 2008, 04:11:59 pm
Image size:	3072 × 2304
Image DPI:	300 pixels/inch
Color model:	RGB
ColorSync profile:	sRGB IEC61966-2.1
Exif and Image Description	
Color Space:	sRGB
Components Configuration:	1, 2, 3, 0
Compressed Bits Per Pixel:	4
Contrast:	Normal
Custom Rendered:	Normal process
Digital Zoom Ratio:	1
Exif Version:	2.2
Exposure Bias Value:	0
Exposure Mode:	Auto exposure
Exposure Program:	Normal program
Exposure Time:	1/13
File Source:	DSC
Flash:	Off, did not fire
FlashPix Version:	1.0
FNumber:	3.3
Focal Length:	6.3
Focal Length In 35mm Film:	38
Gain Control:	Low gain up
ISO Speed Ratings:	200
Light Source:	unknown
Max Aperture Value:	3.4
Metering Mode:	Pattern
Pixel X Dimension:	3,072
Pixel Y Dimension:	2,304
Saturation:	Normal
Scene Capture Type:	Standard
Scene Type:	A directly photographed image
Sharpness:	Normal
Subject Distance Range:	unknown
White Balance:	Auto white balance
Make:	NIKON
Model:	COOLPIX S50
Orientation:	1 (Normal)
Software:	Adobe Photoshop CS2 Macintosh

Figure 5.1. A Representation of Automatically Generated Technical Metadata for a Digital Photograph.

- Identifiers
- Structural types

- File descriptions
- Sizes
- Properties
- Software and hardware environments
- Source information
- Object history
- Transformation history
- Context information
- Digital signatures
- Checksums

In 2015 the third version of the *PREMIS Data Dictionary for Preservation Metadata* was published; it is the standard for preservation metadata used in the United States.

Rights and Access Metadata

Rights and access metadata is information about who has access to information resources, who may use them, and for what purposes. It deals with issues of creators' intellectual property rights and the legal agreements allowing users to access this information. It addresses questions such as: Who can access an information resource and for what purposes? Who can make copies? Who owns the material? Are there different categories of information objects in the collection? Are there different categories of users who can access different combinations of those objects?

In rights and access metadata, information about parties, contents, and transactions can be found. Some typical rights metadata elements include the following:

- Access categories
- Identifiers
- Names of creators
- Names of rights holders
- Dates of creation
- Copyright status
- Terms and conditions
- Access restrictions
- Periods of availability
- Usage information
- Payment options

An example of a rights metadata schema is the Open Digital Rights Language schema.[16]

Meta-Metadata

Not only can metadata track administrative data about a resource, but it can also track administrative information about its metadata. So, if metadata is "data about data," then the "data about the data about data" is *meta-metadata.* Meta-metadata is important for ensuring the authenticity of the metadata and tracking internal processes. While some meta-metadata resides within some types of descriptive records (e.g., the record creation information and modification dates in a MARC record), other meta-metadata must be tracked in other ways. As is the case with general administration metadata, few schemas for meta-metadata exist, but in 2003 the Dublin Core Metadata Initiative's (DCMI) Administrative Metadata Working Group approved the "AC-Administrative Components,"[17] a metadata element set to describe and manage metadata.

This schema created by Hytte Hansen and Leif Andresen is loosely based on an administrative metadata element set proposed four years earlier. AC focuses on elements that describe attributes of the metadata record, changes and updates, and information for the interchange of records. Its elements include, among others,

- Handling
- Action
- Database
- Contact
- Affiliation
- Transmitter

At this writing, however, there is no visible evidence that the schema has been implemented widely.

Structural Metadata

If an information institution receives a digital object, one that it had no part in creating, could the object be opened? If the digital object were a complex, multi-file entity, would the pieces fit together? Would patrons be able to use the resource? Without structural metadata—probably not.

Structural metadata refers to the makeup or internal arrangement of the digital object, dataset, or other information resource that is being described. It is the data needed to ensure that a digital resource functions properly and can be used and navigated by the patron. It refers to how individual related files are bound together to create a working digital object, how the object can be displayed on a variety of systems, and how it can be stored and disseminated. It deals with what the resource is, what it does, and how it works. Structural metadata captures the following kinds of information:

- Document types and their structures
- File types
- Object behaviors or functionality
- Associated search protocols
- Hierarchical relationships
- Sequencing and grouping of files
- Parent objects
- Paging information
- Associated files

Structural metadata can be included in the headers or bodies of some types of electronic documents, but in most metadata schemas, structural elements are not well represented. For most schemas—which focus on descriptive metadata—it is simply unnecessary or inappropriate for the type of resource being represented. For example, with a single digital image, extensive structural metadata is not needed, but once that image becomes part of a complex multimedia object that must operate as a single resource, structural metadata becomes necessary. Sometimes structural metadata is referred to as *display metadata,* and sometimes it is mistaken for technical metadata (discussed in the administrative metadata section above).

The use of structural metadata is not new, but the terminology used to describe it is more recent. An early, successful implementation of structural metadata is the *page-turner model.* A page-turner is used for materials with a defined internal structure and content that is meant to be viewed in a prescribed sequence (e.g., an e-book). It provides structure for the contents to be displayed and for the user to navigate through the information resource as one normally pages through a book. The page-turner may allow the user to navigate through the resource on more than one level, that is, at a chapter level and at a page level. The page-turner uses structural metadata to bind together individual images of pages to form a complete object (again, this may be on the level of an e-book, a volume of a set, or a chapter). It may also use structural metadata to associate a text file with each image of individual pages, so that the intellectual contents of the page image are searchable. Structural metadata can also associate these images with thumbnail images of the pages, images of greater or lesser resolution, HTML or XML-formatted pages associated with the web interface (though this may be done on-the-fly), a separate image file (or files) for an illustration on the page, or some other file or link. Some other early examples of structural metadata schemas include the Electronic Binding DTD (Ebind)[18] and the Making of America II project (MOA2),[19] both of which have since been replaced by METS.

METS (Metadata Encoding & Transmission Standard)

One of the best examples of structural metadata is found in the Metadata Encoding & Transmission Standard (METS), developed as an initiative of the Digital Library Federation and maintained by the Library of Congress Network Development and MARC Standards Office.[20] METS is an XML schema for encoding structurally complex digital objects into a single document that includes descriptive, administrative, and structural metadata. It provides metadata aimed at managing, preserving, displaying, and exchanging digital objects in a digital library environment. Although descriptive and administrative metadata are included in METS documents, the primary focus of the schema is on the structure of digital objects. It is extensible and modular in its approach.

A METS document comprises seven components.

- METS header
- Descriptive metadata
- Administrative metadata
- File section
- Structural map
- Structural links
- Behavior metadata

The *METS header* contains meta-metadata such as the creator and the creation date of the METS document. For the sections on *descriptive metadata* and *administrative metadata*, METS does not define the elements to be included. For descriptive metadata, it allows the creators to choose from a number of extension schemas, including:

- **MARC Bibliographic Format**: standard bibliographic metadata found in library catalogs
- **MODS (Metadata Object Description Schema)**: an XML-encoded adaptation of MARC
- **EAD**: an encoded record format for archival finding aids
- **Dublin Core (DC)**: a general descriptive metadata schema primarily used with online resources
- **VRA (Visual Resource Association) Core**: a specialized metadata schema used to describe art and images
- **TEI (Text Encoding Initiative) Header**: metadata used to describe textual and literary works
- **Data Documentation Initiative (DDI)**: an international schema used to describe statistical and social science data

The administrative metadata section can be divided into four subgroupings: technical, intellectual property rights, source, and digital provenance metadata. For these areas, PREMIS, AudioMD Schema, VideoMD Schema, and MIX have been endorsed as extension schemas that may be used to complete the section. Both administrative and descriptive metadata can be held internally or externally. In other words, the metadata can be included inside a METS document or the METS object can point to a separate metadata record outside of the METS document. Both approaches can be used within the same METS document.

The next component, the *file section*, is an inventory of all the files used to create the digital object. Files may be divided into hierarchically subordinate groups. For example, in a digital object that consists of images accompanied by text, there might be file groups for the thumbnail images, for the master images, for TEI encoded text, and for PDF versions. The *structural map* specifies the ways in which all the files fit together to create the digital object. The map is hierarchically structured and allows the user to navigate from one part of the digital object to another (e.g., from track one to track five on a digitized sound recording). The *structural links* section keeps a record of the hyperlinks and lateral relationships among individual files in the structural map. The *behavior section* describes how the object is to function or perform. As of 2017, METS 1.11 is the current version of the XML schema, but another version of METS is being developed to be compatible with RDF, the structure standard underpinning linked data and the Semantic Web. Examples of METS documents can be seen at the Library of Congress's METS website.[21]

METADATA MANAGEMENT TOOLS

As metadata applications have become more common and more schemas flourish, tools and systems have been developed to help deal with this proliferation. Three of these tools and systems are *application profiles*, *metadata registries*, and *crosswalks*.

Application Profiles

As stated earlier, there is no one-size-fits-all metadata schema. Different schemas have been developed for different purposes, different communities, and different materials. All of these schemas have strengths and weaknesses. When looking for a metadata schema to meet the needs of a particular institution, a community, or a particular project, it is sometimes discovered that no one schema satisfies all of the metadata requirements, but that parts of different, existing schemas would work well *if* they could be combined in a new way. A mechanism to allow metadata creators to use various elements from different schemas is the application profile.

An *application profile*, according to Priscilla Caplan, is a formally developed specification that limits and clarifies the use of a metadata schema for a particular user community. "Whether informal guidelines or formal profiles, additional rulesets are generally needed to supplement metadata schemes as published."[22] Application profiles, therefore, are documents that describe a community's recommended best practices for metadata creation. They contain policies, guidelines, and metadata elements drawn from one or more namespaces. A *namespace* is a formal collection of element types and attribute names; it is the authoritative place for information about the metadata schema that is maintained there. The namespace allows metadata elements to be unambiguously identified and used across communities, promoting semantic interoperability.

The elements selected for an application profile may be a small subset of the elements from a single schema, or they may be elements drawn from two or more schemas combined together. An application profile is a formal way to declare which elements from which namespaces are used in a particular application or project or by a particular community. An example is the application profile proposed by the Dublin Core Libraries Working Group (DC-Lib) for describing library materials. It adds two domain-specific elements (*edition* and *location*) from the MODS metadata schema to the basic DC terms and elements.[23]

Application profiles, unlike metadata schemas, do not create or introduce new metadata elements; they simply mix and match existing elements from various schemas, if one schema is not sufficient. Application profiles may also specify what values are permitted by selecting data value standards to be used in certain elements in the record. This is done to instill some consistency in certain fields so that similar resources can be retrieved together. Application profiles may also refine the definitions of elements by using qualifiers to make them semantically narrower or more specific. For example, an application profile may specify that the element *edition* from MODS will be available for use, that a particular list of values should be used for the DC *type* element, and that the DC *date* element may be made more specific with the following element refinements: *available*, *created*, *issued, modified,* and *valid*. Examples of application profiles can be found on the Dublin Core website.[24]

Metadata Registries

Another tool helpful to the metadata creator is the *metadata registry*. A registry is a database used to organize, store, manage, and share metadata schemas. Registries provide information about schemas and their elements, controlled vocabularies, application profiles, definitions, and relationships, using a standard structure as outlined in ISO/IEC 11179–3:2013, "Information Technology–Metadata Registries (MDR)—

Part 3: Registry Metamodel and Basic Attributes."[25] Metadata registries, once they become more widely implemented, will help greatly to improve interoperability among metadata schemas. A problem is that few registries exist at this time. It is expected that in the future they will be used to exchange information, clarify meaning and usage, and prevent duplication of effort. It is hoped that metadata registries may be a source of machine-understandable information about schemas to support activities and agents online. A widely known example is the Open Metadata Registry.[26]

Crosswalks

Crosswalks, too, are tools used to achieve interoperability, specifically semantic interoperability. Without that one-size-fits-all schema, crosswalks are needed so that users and creators understand equivalence relationships among metadata elements in different communities. Crosswalks help us see, for example, that the 100 field in a MARC record for the primary (or first named) creator is roughly equivalent to the *Creator* field in a Dublin Core record. According to Margaret St. Pierre and William LaPlant, "A crosswalk is a specification for mapping one metadata standard to another. Crosswalks provide the ability to make the contents of elements defined in one metadata standard available to communities using related metadata standards."[27] They go on to observe that creation and maintenance of a crosswalk is difficult and susceptible to error. One needs expertise in each metadata schema included in the crosswalk. But because each standard is developed by experts in a particular field with inherent specialized terminology, persons with expertise in several standards are rare. A key difficulty in creating and maintaining crosswalks is the element-mapping process. The mapping might not be too difficult if the schemas are relatively simple, are for the same types of materials, or have many overlapping concepts. It is far more complex, however, when the mapping is cross-domain, is between schemas of different levels of complexity, or is between schemas with great semantic differences. Generally, the more metadata schemas that are included in the crosswalk, the more difficult it is to do the mapping.

It is important to remember that the conversion process, via crosswalks, from one schema to another lacks precision. Few metadata crosswalks provide round-trip conversion with no loss of data. Some data are lost in one direction or the other (or both). However, until technology that allows machines to understand and link the meanings of various metadata elements is developed further, crosswalks are still useful tools for semantic interoperability.

The following is a small portion of a crosswalk—*Dublin Core Metadata Element Set Mapping to MODS Version 3.0*—which can be found in full on the LC website.[28] In each row, equivalent elements are provided. These are suggestions as to which MODS element may be used when converting from a Dublin Core record to MODS.

Dublin Core	MODS
Description	**<abstract>**
	<note>
	<tableOfContents>
Date	**<dateIssued>**
	<dateCreated>
	<dateCaptured>
	<dateOther>
Rights	**<accessCondition>**

One can see that in some cases one element of a schema is equivalent to *more* than one element in another. For others, though, there may be a one-to-one correlation, or there may be no correlation. It depends on the elements and the schemas involved.

Other Tools

Other tools currently in use include harvesting technologies, which involve automated agents that go out on the web to harvest metadata at a minimal level of complexity. This information is stored and retrieved as needed. The most widely known harvesting project is the Open Archives Initiative Protocol for Metadata Harvesting (OAI–PMH) in which data providers expose their metadata in the form of Dublin Core-based XML documents for collection by harvesters.[29] This metadata, once collected, can then be added to large, searchable metadata indexes. In addition, metadata creation tools, software, and templates (i.e., forms with blank fields that serve as patterns for different metadata schemas) are among the instruments available to improve productivity and consistency in metadata creation.

METADATA MODELS

Around the time of the new millennium, as more and more metadata schemas were being developed, there was interest in developing a shared vision of metadata to help unify diverse information communities in their approaches. There were concerns that the various metadata schemas that were being developed would not be interoperable and that some sort of overarching model or set of policies was needed in order to make resources discoverable on a wide scale. This notion was never fully realized—there is no single metadata schema or policy manual or conceptual model that satisfies the needs of *all* metadata creators, communities, and stakeholders. As time has passed, thoughts about our approaches to metadata have changed. There is a better understanding now that not every community needs to be grounded in a single conceptual model or schema. Instead, the focus is on how to share our individualized, community-specific metadata efficiently and effectively with little loss of data. There is continuing interest, however, in a shared understanding of the process of resource description *within* each of the various information communities. In order to achieve this common vision, individual information communities have designed their own metadata models. For example, there is a model specifically created to address the universe of bibliographic resources, one for cultural heritage resources in museums, and so on. In addition, the work being done to develop the Semantic Web has led to the creation of a data structure model that is key to sharing metadata of all types across communities.

So, what exactly is a *model*? *FOLDOC* states that it is "a description of observed or predicted behaviour of some system, simplified by ignoring certain details. Models allow complex systems, both existent and merely specified, to be understood and their behaviour predicted."[30] Merriam-Webster defines *model* as "a description or analogy used to help visualize something . . . that cannot be directly observed."[31] Models may be used for several reasons, including specifying a broad conceptual framework for the entire bibliographic universe or prescribing specific structural requirements for metadata statements. In this section, we discuss five metadata models:

- the "Functional Requirements" family of standards, including *Functional Requirements for Bibliographic Records* (FRBR) and related models created by the International Federation of Library Associations and Institutions (IFLA), including the more recently developed *IFLA Library Reference Model* (LRM);
- Records in Contexts (RiC), the conceptual model currently being developed by the International Council on Archives' Expert Group on Archival Description (EGAD);
- the Conceptual Reference Model (CRM) developed by the International Committee for Documentation (CIDOC) of the International Council of Museums;
- the DCMI Abstract Model (DCAM) from the Dublin Core Metadata Initiative (DCMI); and
- the Resource Description Framework (RDF), the data-structure model established by the World Wide Web Consortium (W3C) to support data exchange.

IFLA's "Functional Requirements" Family of Standards

In 1998 IFLA published *Functional Requirements for Bibliographic Records*.[32] This report and the conceptual model it details are often referred to as FRBR, which is pronounced *Ferber* by some and as the initials *F.R.B.R.* by others. The report is an attempt to apply an *entity-relationship model* (E-R model)[33] to the various components of the bibliographic universe. (For a brief overview of E-R models, see Textbox 5.1.) It is important to point out that FRBR is an abstract conceptual model; it is not a system design, a metadata schema, a record structure, a content standard, or an encoding format. It is a model that enumerates the entities found in the bibliographic universe and gathers those entities into groups based on the functions or roles of those entities. FRBR identifies the attributes or characteristics associated with each entity, as well as relationships that exist among and within the entity groups. In addition, the conceptual model identifies several *user tasks*—tasks that users perform while using information retrieval tools and interacting with bibliographic data. These tasks inform the entire conceptual model, helping readers to remember the reasons why the bibliographic universe is organized. The model maps each entity's attributes and its relationships to the user tasks, which allows catalogers to see the significance of each element in a bibliographic description. This conceptual model helps answer some important questions about what metadata might be of the most value to library users and how that metadata is used. Finally, FRBR provides the international cataloging community with recommendations regarding a minimal set of required elements that are necessary to include in resource descriptions to support the user tasks. Ultimately, the FRBR model identifies four user tasks, three groups of entities, and myriad attributes and relationships among the entities. Some of these are addressed in the sections that follow.

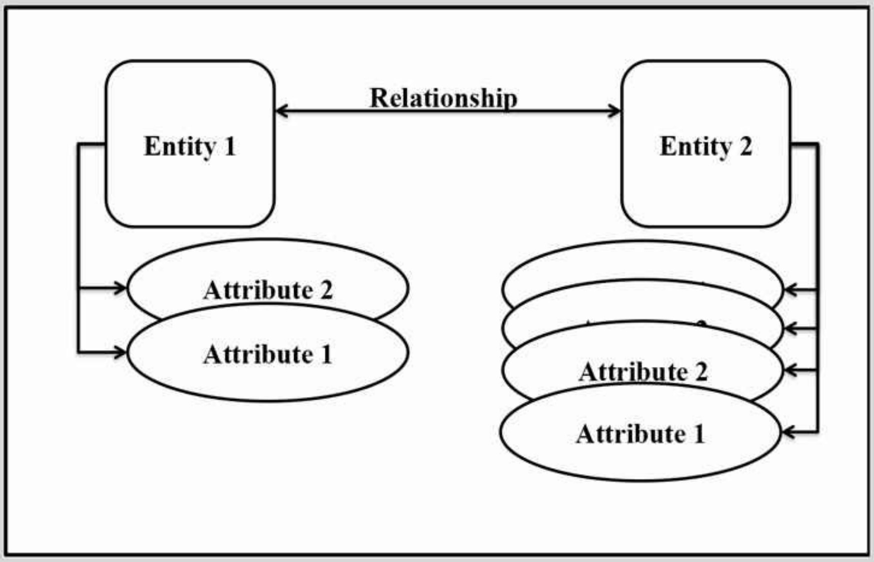
In the years since FRBR's publication, IFLA has produced additional related reports. These include

- *Functional Requirements for Authority Data* (FRAD) in 2009,[34]
- *Functional Requirements for Subject Authority Data* (FRSAD) in 2010,[35] and
- *IFLA Library Reference Model* (LRM) in 2017—an attempt to harmonize the three earlier "Functional Requirements" models (originally titled *FRBR-Library Reference Model* or FRBR-LRM).[36]

FRAD, like FRBR, is an attempt at modeling the bibliographic universe, but instead of focusing on bibliographic records FRAD is concerned with authority data and the concepts related to authority control. It is an extension of the FRBR model, adding more entities, attributes, relationships, and user tasks. For more information about authority control and FRAD, please see Chapter 8. FRSAD is a high-level conceptual model of the subject relationships existing in the bibliographic universe. It identifies additional entities, relationships, and user tasks. At the time of this writing, LRM is under review, but it is expected for IFLA to approve the standard in 2017. Its major changes are described at the end of this section on FRBR. For a more in-depth discussion of FRBR, FRAD, and FRSAD, please see the 11th edition of *Introduction to Cataloging and Classification* by Daniel N. Joudrey, Arlene G. Taylor, and David P. Miller.[37]

A Brief Overview of Entity-Relationship (E-R) Models

In E-R models, there are three classes of concern; they comprise *entities* and *relationships* obviously, but *attributes* are also included in the basic structure. A very simple diagram of an E-R model is presented here:



Entities are things. They are key objects of interest about which information is sought. Entities are capable of being distinctly identified. A person, a creative work, a corporate body, an event, and a concept are all examples of entities. An entity is often a focal point for gathering data.

Relationships are reciprocal associations or connections between entities. These connections are usually expressed in a way to indicate that the relationship flows in each direction between the entities. An entity may have one or more relationships, but only relationships of significance are usually documented. Among many others, examples of relationships include creation relationships, whole–part relationships, appellation relationships, and so on.

Attributes are characteristics of entities or relationships. Attributes are identifying features that help in understanding the essence of what is being described. Examples of attributes may include the title of a work, the birth date of a person, publication date, and so on.

To provide a loose analogy, one could equate the components of E-R models to basic grammar. One could look at entities as nouns, relationships as verbs, and attributes as adjectives and adverbs. This is not an exact correlation, however, because conceptual models may apply the elements slightly differently.

Textbox 5.1. A Brief Overview of E-R Models.

User Tasks

One of the primary purposes of creating metadata is to help users find the information resources they might need. It is advantageous then, when creating metadata, to look at the objectives that users may have when approaching an information retrieval system so that their goals and needs may be met. The following is FRBR's list of user tasks. Upon examination, it becomes clear that these tasks are related closely to both Cutter's objectives of the catalog[38] and the Paris Principles mentioned in Chapter 2.[39] FRBR states that users approach an information system in order to accomplish the following activities:[40]

- **To Find**: Users approach retrieval systems to search for information resources that meet certain criteria. They may wish to find the articles published by *Cataloging & Classification Quarterly* in 2017, books about Minnie Earl Sears, or DVDs of Sondheim musicals.

- **To Identify**: The metadata records found in retrieval tools help users to recognize the appropriate entities and information resources. This may involve distinguishing among similar entities or identifying an item that corresponds to a citation in a bibliography. For example, a user may wish to find the works by Michael Gorman on cataloging. The user will depend upon the system to (1) collocate all of the works by each author named Michael Gorman, and (2) distinguish the Michael Gorman who wrote about cataloging from the Michael Gorman who wrote about religious views on abortion, the Irish fiddler named Michael Gorman, and the Michael Gorman who wrote about German history.
- **To Select**: Systems help users to choose information resources that are appropriate for their needs. This may involve attributes such as content, format, language, edition, or system requirements. For example, a user in Spain may need a copy of *Hamlet* in Catalan, but not in Spanish, Basque, Galician, Aranese, or the original English.
- **To Obtain**: Users approach systems in order to acquire or gain access to information resources. Users depend on the system to provide call numbers, the names of journals containing articles sought, or URLs for direct access to resources.

In her book, *The Intellectual Foundation of Information Organization*, Elaine Svenonius offers a fifth user task:

- **To Navigate**: This objective takes into account the information-seeking behavior of some users who may not be able to articulate completely their information needs, but will instead use the structure of the information system to find the information they are seeking.[41]

The navigation task depends heavily on the information system being able to identify and to take advantage of the relationships that exist among information resources; this can only happen if those relationships are represented in metadata descriptions.

Reflecting its orientation toward authority data, FRAD includes another type of user in its model: *metadata creators.* This addition supplements the general users of bibliographic data (e.g., patrons in a library) addressed in FRBR. Consequently, the user tasks identified in FRAD are somewhat different. The first two tasks are taken directly from FRBR (i.e., *to find* and *to identify*). The second two, however, are quite different. These are *to contextualize* and *to justify,* tasks performed by those creating authority data. According to FRAD, *contextualization* is to "place a person, corporate body, work, etc., in context; clarify the relationship between two or more persons, corporate bodies, works, etc.; or clarify the relationship between a person, corporate body, etc., and a name by which that person, corporate body, etc., is known."[42] *Justification* refers to documenting acceptable reasons for the choice of the forms of names or titles as the preferred forms.

The FRSAD model has yet another variation on its users and user tasks. FRSAD recognizes four sets of users:

- general end-users (similar to those emphasized in FRBR);
- information professionals who create and maintain metadata (similar to those added by FRAD);
- those who create and maintain subject authority data (i.e., catalogers and the creators of controlled vocabularies); and
- reference librarians and other professionals who search on behalf of general end-users.[43]

FRSAD includes the first three user tasks from FRBR (i.e., *to find*, *to identify*, and *to select*); its fourth task is *to explore*. This refers to exploring "relationships among terms during cataloguing and metadata creation," as well as exploring "relationships while searching for bibliographic resources."[44] Different versions of the lists of user tasks are presented side-by-side in Table 5.3. In addition to FRBR, FRAD, and FRSAD, the table includes a version found in another IFLA document, *Statement of International Cataloguing Principles* (ICP), in which they are referred to as "objectives and functions of the catalogue," rather than as *user tasks*.[45]

Table 5.3. User Tasks Found in Different IFLA Models.

FRBR 1998	FRAD 2009	International Cataloguing Principles 2009	FRSAD 2010
Find	**Find**	**Find**	**Find**
Identify	**Identify**	**Identify**	**Identify**
Select	**Contextualize**	**Select**	**Select**
Obtain	**Justify**	**Obtain**	**Explore**

In order to meet the needs of all users, metadata should be created with the users and their tasks in mind. The metadata created should include descriptions of the various entities of interest in the bibliographic universe, their attributes, and the relationships among those entities. Every piece of the description should be useful in meeting one of the user tasks outlined above. If an attribute or relationship does not meet one of the user tasks, metadata creators should ask whether or not that information is needed.

FRBR Entities and Attributes

In the original FRBR model, entities are divided into three groups and are identified based on their role or function. The first group contains four entities—*work*, *expression*, *manifestation*, and *item* (WEMI)—that are the products or results of "intellectual or artistic endeavour"[46] that are named or described in bibliographic records. Group 1 comprises the components of a bibliographic resource; it represents what is collected in libraries and other information institutions. The four entities are connected by high-level relationships, as can be seen in Figure 5.2.[47] Each of the Group 1 entities is explained in more detail below, as are the entities from the other two groups.

The first entity, *work,* is the top, most general level in Group 1. It is an abstract concept that essentially exists only in the mind of the creator; there is no single physical object associated with it. It can be described as the distinct idea or the artistic or intellectual vision of the creator. It is recognized through individual expressions of the work. Examples of works include Shakespeare's *Romeo and Juliet,* Mozart's *The Magic Flute,* and Michelangelo's *David.*

The second level, *expression,* is a realization of a work in the form of text, numbers, musical notation, choreographic symbols, sounds, movement, images, and so on. It indicates *how* the content of the work is being communicated. For example, the expression level encompasses "the specific words, sentences, paragraphs, etc. that result from the realization of a work in the form of a text."[48] It is "the specific intellectual or artistic form that a *work* takes each time it is 'realized.'"[49] There can be more than one expression of a work. For example, a work might be expressed originally in the form of French text, but it may also have a Somali translation, a translation in American Sign Language, a spoken-word performance, a version expressed through French text but with accompanying illustrations, and so on. Expression is also an abstract, intangible content-related entity with no particular physical carrier (i.e., you cannot touch or buy a textual expression; only the physical item that contains it). Imagine the words of a book floating in the air before they are manifested on a computer screen or on a piece of paper—that would be an expression. Expressions are reflected in updated editions, new or revised versions, as well as any translations, of a resource.

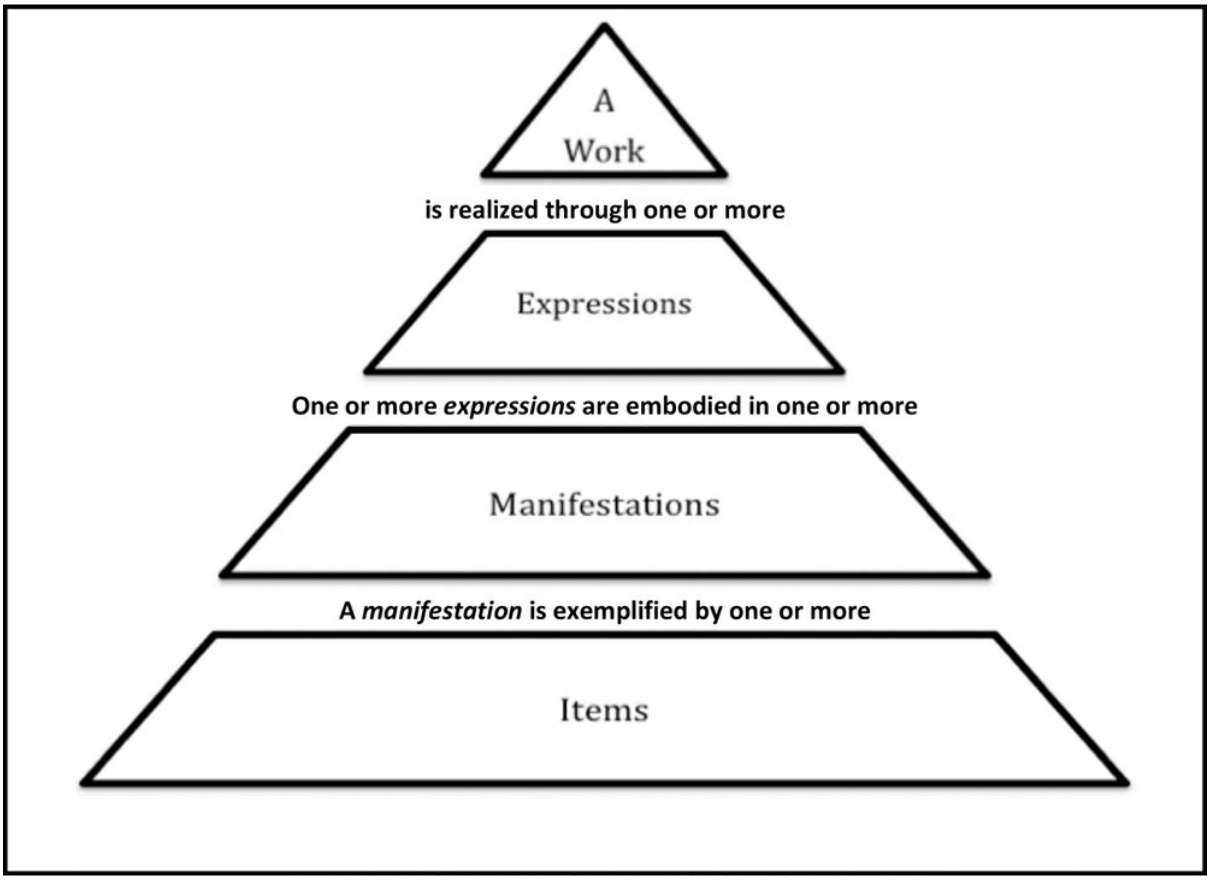


Figure 5.2. Group 1 Entities in the FRBR Model.

In order for a work and its expression to take material form, it must be embodied in the third entity, *manifestation.* Manifestations are sets of physical resources that contain the work and its expression(s). For example, when a book is published, we have a set of printed text manifestations (although *physical* should be interpreted loosely here, as it might consist of electronic impulses in a computer, for example). It is the situation in which the same content is reproduced, even though the format may be different. When a musical work is expressed in a studio performance, it can be recorded and then transferred onto compact discs, vinyl, magnetic tape, electronic sound files, and so on; each of these carriers (i.e., physical or electronic instantiations) represents a different manifestation.

The fourth level, *item,* represents a single exemplar or instance of a manifestation. The item is the individual copy of an information resource (e.g., my copy of this textbook, your set of *The Walking Dead: Season 6* DVDs). As a rule, exemplars are identical to each other, but occasionally they can be different in interesting ways. For example, there might be a damaged copy, a copy autographed by the author, or a copy bound by a library's rebinding department. The utility of the item level is to help distinguish one exemplar from another exemplar of a manifestation.

Each of the FRBR Group 1 entities has its own set of attributes associated with it. Attributes help users to find, identify, select, and obtain information resources in retrieval tools. A brief sample of attributes follows:

- **Attributes of Works**: Title of Work, Date of Work, Intended Termination, Intended Audience, Context for the Work, Other Distinguishing Characteristic
- **Attributes of Expressions**: Form of Expression, Date of Expression, Language of Expression, Summarization of Content, Extensibility of Expression, Other Distinguishing Characteristic
- **Attributes of Manifestations**: Title of the Manifestation, Statement of Responsibility, Edition/Issue Designation, Place of Publication, Publisher, Date of Publication
- **Attributes of Items**: Item Identifier, Fingerprint, Provenance of the Item, Marks/Inscriptions, Exhibition History, Condition of the Item

Although some of the attributes may seem repetitive, the values of those attributes may not be identical among the four Group 1 entities. For example, the title of a work and the title of a manifestation may be the same or they may be different based on the language of the expression, title changes over time, the inclusion or exclusion of introductory words, or other variations.

The second group contains those entities responsible for the creation, production, dissemination, and ownership of the Group 1 entities. In this group, the original FRBR model includes only *persons* and *corporate bodies;* a third entity, *families,* was added in FRAD. The FRBR and FRAD models establish that persons, families, and corporate bodies create works, realize expressions, produce manifestations, and own or provide access to items. This framework reflects the entity-relationship model well. The concept of *creator,* an element typically found in most metadata schemas, is *not* an attribute of a work, but instead is recognized as a separate entity (a person, family, or corporate body) that is connected to a work through a specific kind of relationship (i.e., a *created by/creates* relationship).

FRBR Group 3 entities are those that can be the subjects of works. These include: *concepts, objects, events,* and *places.* Concepts are abstract ideas, objects are physical things, events are occurrences, and places are locations. Group 3 also includes all the other entities in the FRBR model, because works can be about other works, expressions, manifestations, and even individual items, as well as about persons, families, and corporate bodies.

In addition to the three groups of entities identified in FRBR, FRAD has identified five additional entities specifically related to authority data. These entities include *name, identifier, controlled access point, rules,* and *agency.* Like the FRBR entities, the FRAD entities have attributes and relationships associated with each of them. FRSAD has introduced two additional entities in its highly abstract, greatly generalized model of subject relationships: *thema* and *nomen*. *Thema* refers to any type of subject matter (e.g., a topic, a time period, a work), and *nomen* refers to that thing's name or label. In short the model states that a work is about *something*, and that *something* has a name. The FRSAD model also identifies attributes and relationships for *thema* and *nomen*.

FRBR Relationships

The final major component in an entity-relationship model is the existence of relationships among the various entities. FRBR and its related models identify some widely applicable, general relationships among entities, as well as some specific types of bibliographic relationships that can exist. General relationships include:

- the inherent relationships among the Group 1 entities: work, expression, manifestation, and item (i.e., a work *is realized through* one or more expressions, which *are embodied in* one or more manifestations, which *are exemplified in* one or more items);
- the relationships between Group 1 entities and Group 2 entities: persons, families, and corporate bodies (i.e., *created by, realized by, produced by,* and *owned by* relationships); and
- the subject relationships between works and the entities in groups 1–3.

The more specific types of bibliographic relationships represented in FRBR have been influenced by Barbara Tillett's taxonomy of bibliographic relationships[50] and Richard Smiraglia's research focusing on derivative bibliographic relationships[51] (see discussion in Chapter 8). Some of the relationships identified in the original FRBR model are listed in Table 5.4.

Table 5.4. Selected Relationships in the FRBR Model.

Type	Explanation	Sub-types	Examples
Work-to-Work Relationships	Relationships that exist between two separate works	Successor, supplement, complement, summarization, adaptation, transformation, and imitation relationships	Sequels, concordances, indexes, librettos, digests, paraphrases, parodies, dramatizations, etc.
Whole–Part Relationships	These are relationships that have some sort of component aspect	Independent or dependent whole–part relationships	Relationship between a chapter and the whole work, between a book and the series of which it is a part, between an article and the journal in which it was published, etc. Whole–part relationships may be found at the work,

			expression, manifestation, and item levels.
Expression-to-Expression Relationships	**Relationships that exist between two separate expressions**	**Relationship between two expressions of the same work**	***Same work:* Abridgements, revisions, translations, arrangements, etc.**
		***Or* between expressions of different works**	***Different works:* identical to the work-to-work relationships listed above.**
Manifestation-to-Manifestation Relationships	**Relationships that exist between two separate manifestations**	**Reproduction and alternate relationships**	**Reprints, microfilms, and facsimiles; simultaneously published editions; multiple formats; etc.**
Item-to-Item Relationships	**Relationships that exist between two separate items**	**Reproduction and reconfiguration relationships**	**Binding changes, extracts, etc.**

It is a brief list; for further information about relationships in FRBR, please see the model itself.[52] In addition to those relationships enumerated in the FRBR model, FRAD and FRSAD identify several additional types of relationships that are applicable to their specific purposes.

FRBR and its related models have played pivotal roles in IFLA's development of the latest set of international cataloging principles.[53] These principles and the FRBR and FRAD models have also been influential in the development of *RDA: Resource Description & Access,* the content standard used to guide contemporary descriptive cataloging activities in the United States, Canada, Great Britain, Germany, and many other places around the world.[54]

IFLA Library Reference Model

The Need for Harmonization. In 2009, when FRAD was first published, readers quickly noticed there were differences between it and the original FRBR model published more than 10 years earlier. This was not surprising, considering the time that had passed and the different foci of the reports (i.e., bibliographic versus authority data), but the discrepancies were still of concern to many. The following are some of the major changes that were introduced in FRAD:

- New entities were included (e.g., families, names, rules).
- Half of the user tasks were different.
- A new type of user was considered in the model.
- New types of relationships were introduced (e.g., person-to-person relationships, family-to-corporate body relationships).
- The specific bibliographic relationships included were based on a different conceptual framework.

Then later, when FRSAD was published, more incongruities appeared. FRSAD increased the types of users considered, it modified the user tasks, but most significantly, it disregarded both FRBR's and FRAD's approaches to subjects in favor of a much more abstract, higher-level conceptual framework. These discrepancies were starting to create tensions in the library cataloging community. When the three conceptual models underpinning contemporary cataloging disagree on something as fundamental as subject access, something must be done. The reasons for creating the IFLA Library Reference Model are well summarized in its background statement.

> Inevitably the three FR models, although all created in an entity-relationship modelling framework, adopted different points of view and differing solutions for common issues. Even though all three models are needed in a complete bibliographic system, attempting to adopt the three models in a single system required solving complex issues in an ad hoc manner with little guidance from the models. Even as FRAD and FRSAD were being finalized in 2009 and 2010, it became clear that it would be necessary to combine or consolidate the FR family into a single coherent model to clarify the understanding of the overall model and remove barriers to its adoption.[55]

Work on LRM began in 2010, the same year that FRSAD was published. As was the case with the original FRBR model, LRM is a high-level conceptual model; it is not a content standard; it does not provide specific rules or guidelines. It is not only a consolidation (or reconciliation or harmonization) of the three models, it is also a broadening or generalization of the overall conceptual model. It has been influenced by the conceptual reference model developed by the International Committee for Documentation of the International Council of Museums (see more on this model below). It has eliminated some administrative focus and some of the more specific details found in the three models. At the time of this writing, LRM is in final draft form. When it is adopted, it is to supersede the previous three models.

Changes in the LRM. Despite this section's title, we will start by looking at what has *not* changed. The proposed LRM is an E-R model like its predecessors. It defines a user population of interest and a set of user tasks. The focus of much of the model is on key entities, attributes, and relationships found in the bibliographic universe. It uses a structure and vocabulary similar to those found in the three previous FR models.

In LRM, however, the user population is reconsidered and simplified. It is "primarily concerned with the data and functionality required by end-users (and intermediaries working on behalf of end-users) to meet their information needs."[56] It has returned, in some ways, to the relative simplicity found in the original FRBR document. While the consolidated model *may* address the needs of information professionals in some respects, they are not the user group of primary concern. Due to this altered view of the user population, the user tasks were reviewed, resulting in the elimination of *justify* because it was oriented to administrative tasks. Some aspects of FRAD's *contextualize*, which has been removed as well, have been incorporated into the task *explore*. The user tasks also do not include *navigate*, which was proposed by Svenonius and included in IFLA's ICP. These changes result in a somewhat new amalgamation of user tasks in LRM:

- **Find**: To bring together information about one or more resources of interest by searching on any relevant criteria;
- **Identify**: To clearly understand the nature of the resources found and to distinguish between similar resources;
- **Select**: To determine the suitability of the resources found, and to be enabled to either accept or reject specific resources;
- **Obtain**: To access the content of the resource; and
- **Explore**: To discover resources using the relationships between them and thus place the resources in a context.[57]

The entities in LRM have also changed. The total number of entities has been reduced, and they are now structured into three hierarchical tiers (see Figure 5.3). This allows "the transfer of attributes and relationships from the superclass to its subclasses."[58] One of the more noticeable changes is that the top tier of the model is populated by only a single entity. It is based on and generalized from the FRSAD entity, *thema*, but it has been renamed with the Latin word *res* (i.e., *thing*). Everything in the model, therefore, is a res, but each thing gets more specific at the lower tiers of the model. Tier two includes the former Group 1 entities (work, expression, manifestation, item), though these too are slightly redefined (e.g., to clarify that a manifestation is a set of items in the same particular form).[59] The former Group 2 entities (person, family, corporate body) are now subsumed under the new second-tier entity, *agent*. Agent is a superclass of two third-tier entities: *person* and *collective agent*. *Family* and *corporate body* are no longer considered entities as such, but they may be identified as specific types of collective agents. Both *place* and *time-span* are essentially new entities (thanks to a generalization of the *place* entity from FRBR's Group 3). The final entity, *nomen*, which was originally in FRSAD, has been combined with FRAD's *name* entity because they performed very similar functions (i.e., they identified the labels used for other entities, such as *thema* or *person*). All of the other entities found in the original three models have been deprecated or deemed out of scope for LRM; these include *family*, *corporate body*, *concept*, *object*, *event*, *identifier*, *controlled access point*, *agency*, and *rules*.

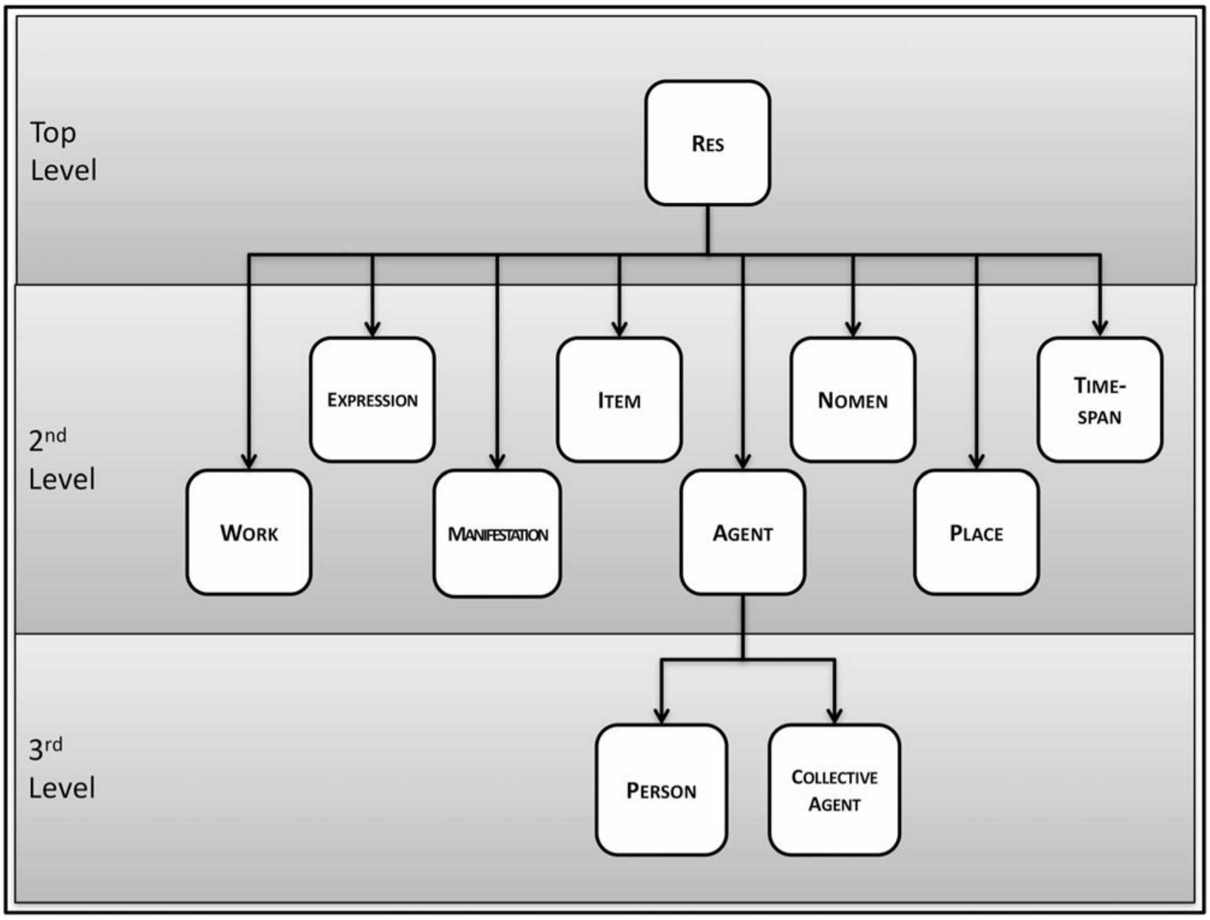


Figure 5.3. Entities in the IFLA Library Reference Model.

In addition to the changes found among the entities, changes are found in the attributes and the relationships in the draft model. Attributes continue to describe entities, and relationships continue to connect entities, but due to the strict hierarchical structure introduced in LRM, both attributes and relationships are now transferable from a superclass to its subclasses. In other words, that which applies to the larger class above (e.g., *agent*) trickles down to entities below it (e.g., *person* and *collective agent*). So, attributes and relationships defined at the highest levels are generally not repeated at lower levels. For example, if *agent* can have a creation relationship to a work, that relationship is also applicable to *person* and *collective agent*. An exception, however, was made for the attribute *category*, which is repeated under a number of entities, due to variations in how that attribute is defined under the different entities.

IFLA LRM includes 37 attributes for its 11 entities. These are briefly listed in Table 5.5. Many of the myriad attributes identified in the earlier models are not listed in the consolidated model. Only essential attributes have been retained, and some new attributes have been introduced.

Table 5.5. Attributes Specified in LRM.

Entity	Attribute	Entity	Attribute
Res	Category	Item	Location
	Note		Use Rights
Work	Category	Agent	Contact Information
	Representative Expression		Field of Activity
Expression	Category		Language
	Extent	Person	Profession/Occupation
	Intended Audience	Nomen	Category
	Use Rights		Nomen String
	Cartographic Scale		Scheme
	Language		Intended Audience

	Key		**Context of Use**
	Medium of Performance		**Reference Source**
Manifestation	**Category of Carrier**		**Language**
	Extent		**Script**
	Intended Audience		**Script Conversion**
	Manifestation Statement	**Place**	**Category**
	Access Conditions		**Location**
	Use Rights	**Time-Span**	**Beginning**
			Ending

Many attributes have been

- removed because they were out of scope or too specific (e.g., *other distinguishing characteristics of work*, *sequencing pattern*, *scheduled treatment*);
- replaced by a relationship to another entity (e.g., the attributes *title of work*, *name of person*, and *manifestation identifier* are now replaced by appellation relationships between a particular entity and a nomen; most attributes related to locations or dates have been replaced with relationships to a *place* or *time-span* entity);
- merged into other similarly focused attributes (e.g., *statement of responsibility* and *edition* have been merged into the new *manifestation statement* attribute); or
- renamed and/or redefined more generically (e.g., *form of expression* is now represented by the attribute *category*).

The attributes listed in LRM under each entity are "representative and are not in any way to be considered an exhaustive listing of attributes that might be determined to be useful in a particular application. An application can define additional attributes to record additional relevant data or to record data at a greater level of granularity than is illustrated."[60]

Relationships, too, have changed. There is a greater emphasis on relationships in this version of the model than was found in the original FRBR report (which had a greater focus on attributes). There are 36 declared relationships. The three major relationships between the WEMI entities remain intact (a *work* is realized in an *expression*, an *expression* is embodied in a *manifestation*, a *manifestation* is exemplified by an *item*), but some of the inherent relationships found in the original FRBR model have been deprecated. The most illustrative example is the change between WEMI and Agents. In the previous model, a person or corporate body has specialized relationships with each entity in Group 1 (e.g., a *person* creates a *work*; a *person* realizes an *expression*; a *corporate body* produces a *manifestation*; a *corporate body* owns an *item*). In LRM, these high-level relationships are both more general and more specific.

- Work was *created by* an Agent
- Expression was *created by* an Agent
- Manifestation was *created by* an Agent
- Manifestation was *distributed by* an Agent
- Manifestation was *manufactured by* an Agent
- Item is *owned by* an Agent
- Item was *modified by* an Agent

Obviously, a *creation* relationship is no longer as narrowly defined as in the original model, where only *works* were created. It now can apply to expressions and manifestations as well. For example, it can now be stated that Lanford Wilson *created* a translation of Chekhov's *Three Sisters*, rather than saying Wilson realized a new English expression of that work.

With the introduction of new entities, it is unsurprising that some additional relationships are needed in the revised model. For the *res* entity, several relationships have been identified:

- *res* is associated with a *res*
- *res* is a subject of a *work*
- *res* has an appellation of *nomen*
- *res* is associated with a *place* (or *time-span*)

Table 5.6. Relationships Specified in LRM.

Focus	Relationships	Focus	Relationships
Res	• ***Res*** **is associated with** ***Res*** • ***Res*** **has association with** ***Place*** • ***Res*** **has association with** ***Time-Span***	**Agents**	• ***Agent*** **is a member of** ***Collective Agent*** • ***Collective Agent*** **has part** ***Collective Agent*** • ***Collective Agent*** **precedes** ***Collective Agent***
High-Level WEMI Relationships	• ***Work*** **is realized through** ***Expression*** • ***Expression*** **is embodied in** ***Manifestation*** • ***Manifestation*** **is exemplified by** ***Item***	**Whole–Part Relationships**	• ***Work*** **has part** ***Work*** • ***Expression*** **has part** ***Expression*** • ***Expression*** **was aggregated by** ***Expression*** • ***Manifestation*** **has part** ***Manifestation*** • ***Place*** **has part** ***Place*** • ***Time-Span*** **has part** ***Time-Span***
WEMI and Agents	• ***Work*** **was created by** ***Agent*** • ***Expression*** **was created by** ***Agent*** • ***Manifestation*** **was created by** ***Agent*** • ***Manifestation*** **was manufactured by** ***Agent*** • ***Manifestation*** **was distributed by** ***Agent*** • ***Item*** **is owned by** ***Agent*** • ***Item*** **was modified by** ***Agent***	**Specific WEMI Relationships**	• ***Work*** **precedes** ***Work*** • ***Work*** **accompanies/complements** ***Work*** • ***Work*** **is inspiration for** ***Work*** • ***Work*** **is a transformation of** ***Work*** • ***Expression*** **is derivation of** ***Expression*** • ***Manifestation*** **has reproduction** ***Manifestation*** • ***Manifestation*** **has alternate** ***Manifestation*** • ***Item*** **has reproduction** ***Manifestation***
Subject and Name Relationships	• ***Work*** **has as subject** ***Res*** • ***Res*** **has appellation** ***Nomen*** • ***Agent*** **assigned** ***Nomen*** • ***Nomen*** **is equivalent to** ***Nomen*** • ***Nomen*** **has part** ***Nomen*** • ***Nomen*** **is derivation of** ***Nomen***		

As was the case in the original FRBR model, the relationships are understood to flow in both directions. Therefore, if an agent created a work, then we also understand that the work was created by the agent.

Unlike FRBR though, IFLA LRM may contain hierarchical relationships (e.g., X IsA Y) or multi-step relationships. For example, to relate a person's pseudonym to the title of a work, several relationships must be identified: (1) WORK was created by an AGENT; (2) AGENT IsA PERSON; (3) AGENT/PERSON IsA RES; (4) RES has an appellation NOMEN1; (5) NOMEN1 is equivalent to NOMEN2. In summary, these five relationships indicate that the creator of a work is a person who is known by more than one name. The list of relationships specified in LRM is found in Table 5.6. More details on specific relationships and the rest of the changes can be found in the LRM documentation.[61]

The family of FR-models has had a transformative effect on our thinking about bibliographic metadata—pushing it toward an entity-relationship model, which naturally points us toward relational database structures. What is interesting is that as the conceptual models have pointed us in that direction, web technologies and the development of Semantic Web tools seem to be reaching out for data management strategies that go far beyond a straightforward relational database. It should also be kept in mind that the FR-family of models was designed to reflect the bibliographic universe. Although that universe, in theory, includes all types of resources, it has been observed that FRBR does not address all types of information resources as well as it does print materials. For example, according to art librarians Murtha Baca and Sherman Clarke, the FRBR model does not apply to many types of resources collected by museums;[62] nor does it easily apply to collections in archives, historical societies, and other similar institutions. That is why other communities have developed their own conceptual models.

Archives: EGAD's Records in Contexts (RiC)

In late 2012 the International Council on Archives (ICA) appointed an Expert Group on Archival Description (EGAD) and charged it with the development of a conceptual model for archival description that would merge four existing archival standards:

- *General International Standard for Archival Description* (ISAD(G)),
- *International Standard Archival Authority Records–Corporate Bodies, Persons, and Families* (ISAAR(CPF)),
- *International Standard Description of Functions* (ISDF), and
- *International Standard Description of Institutions with Archival Holdings* (ISDIAH).[63]

EGAD released a draft conceptual model for Records in Contexts (RiC) in late 2016 for comment. This conceptual model reconciles the four standards and re-conceptualizes archival description by identifying entities, properties of those entities, and relationships those entities can have with other entities. In this way, it is similar to the LRM developments in the bibliographic world. The conceptual model is in its initial stages of development. EGAD ultimately envisions a two-part standard: the conceptual model (RiC-CM) and an ontology (RiC-O). RiC-O will translate the conceptual model expressed in RiC-CM using an encoding standard, the W3C's Web Ontology Language (OWL), to provide the archival community with the ability to use linked data techniques with archival description (for more on ontologies as information organization tools, see Chapter 10). The development of the ontology, however, is dependent on a stable draft of the RiC-CM, so there is much to watch in the near future.

Museums: CIDOC Conceptual Reference Model (CRM)

The International Committee on Documentation (CIDOC) has worked for nearly 20 years on the development of a general data model focusing specifically on improving the exchange of information among museums and other cultural heritage institutions. Within the context of the CIDOC Conceptual Reference Model (CRM), *cultural heritage* is a broad term that refers to the institutions that acquire, preserve, and provide access to resources, bringing together libraries, archives, and museums. In 1999 the first complete edition of the CIDOC CRM was released by the CIDOC Documentation Standards Working Group. CIDOC CRM was subsequently submitted to the International Organization for Standardization (ISO) and was designated an official standard (ISO 21127) in 2006, with a recent update published in 2014.[64]

Although called a *model*, the CRM does not exactly fit the definitions provided earlier in this chapter. It is certainly not a simplified description of a system. There is no single visual representation that can easily communicate the concepts being described. CIDOC CRM is, instead, a highly complex and thoroughly developed *ontology*—a formal representation of the reality of a knowledge domain—comprising hundreds of components. It is intended to promote the efficient and effective exchange of cultural heritage information by providing a common semantic framework to which heterogeneous sources of cultural heritage data can be mapped. It is "intended to be a common language for domain experts and implementers to formulate requirements for information systems and to serve as a guide for good practice of conceptual modelling. In this way, it can provide the 'semantic glue' needed to mediate between different sources of cultural heritage information, such as that published by museums, libraries and archives."[65]

The CRM provides definitions and a formal structure for the description of 94 entities (now referred to as *classes* in CRM) and 168 relationships (referred to as *properties*) used in cultural heritage documentation. Classes are the core components of the CIDOC CRM. A class is a "category of items that share one or more common traits serving as criteria to identify the items belonging to the class," and a property "serves to define a relationship of a specific kind between two classes."[66] Also built into the CRM is a hierarchical understanding of classes and properties, therefore allowing for subclasses, superclasses, subproperties, superproperties, and inheritance relationships among them. Two entities sit at the top of the hierarchy of classes: *CRM Entity* (i.e., a thing) and *Primitive Value* (i.e., a data value), each with several subclasses. *CRM Entity*, for example, contains five subclasses—*Temporal Entity*, *Time-Span*, *Place*, *Dimension*, and *Persistent Item*—some of which have additional subclasses. In some cases, the class hierarchies can get quite deep:

CRM Entity
  Persistent Item
    Thing
      Man-Made Thing

Conceptual Object
Symbolic Object
Appellation
Contact Point
Address

Classes and properties can also be divided into the following 34 *functional units*:[67]

- Acquisition Information
- Appellation Information
- Attribute Assignment
- Changing Thing
- Collection Information
- Condition Information
- De-accession and Disposal Information
- Description Information
- Documentation and References
- Existence Information
- Group Dynamics
- Image Information, Objects and Carriers
- Institution Information
- Location Information
- Mark and Inscription Information
- Material and Technique Information
- Measurement Information
- Object Association Information
- Object Collection Information
- Object Entry Information
- Object Name and Classification Information
- Object Number Information
- Object Production Information
- Object Title Information
- Part and Component Information
- Person Nationality Information
- Planned Activities (design, purpose, use)
- Recorder Information
- Reference Information
- Reproduction Rights Information
- Spatial–Temporal Relationship
- Subject Depicted Information
- Taxonomic Discourse
- Time-Span Information

These categories reflect possible uses for classes and properties within the realm of cultural heritage data. The functional categories are not mutually exclusive. For example, the class *Event* appears in 13 different functional units. Each functional unit has a visual model showing how the classes and properties may be connected.

CIDOC CRM is a complex system designed to represent and share cultural heritage metadata. Although it employs a somewhat different underlying structure (it is based on an object-oriented model rather than on an entity-relationship data model), it can be mapped to other metadata standards, such as the Dublin Core, Encoded Archival Description, and the FRBR model.[68] To further explore this ontology, please see the extensive documentation provided on the CIDOC CRM homepage.[69]

W3C's Resource Description Framework (RDF)

Because no one-size-fits-all metadata schema exists, the metadata community quickly came to realize that multiple descriptions for a single information resource might be created. The Warwick Framework,[70] developed at the Second Dublin Core Metadata Workshop in 1996, is a model for pulling together distinct packages of metadata that are related to the same information resource into a single container. Referred to as *container architecture,* this conceptual model allows different communities to create, maintain, and share their metadata. The Warwick Framework influenced the early development of the Dublin Core, but its greatest impact was as an evolutionary step in the development of the first iteration of RDF, a metadata specification developed by the World Wide Web Consortium (W3C) in 1999, whose basic metadata model helps to promote structural interoperability.[71]

According to the W3C, RDF is a "standard model for data interchange on the Web."[72] It is a "framework for expressing information about resources."[73] It allows us to make statements about *resources*. Resources can be anything (documents, books, people, tangible objects, abstract concepts) found online or in the physical world; in RDF, the term *resource* is synonymous with *entity*.[74] "RDF is intended for situations in which information on the Web needs to be processed by applications, rather than being only displayed to people. RDF provides a common framework for expressing this information so it can be exchanged between applications without loss of meaning. . . . In particular RDF can be used to publish and interlink data on the Web."[75]

RDF enables the exchange and reuse of metadata in ways that are semantically unambiguous; in other words, RDF allows machines to "understand" what is being communicated by the metadata. This is accomplished by structuring all metadata statements in the form of *triples.* An RDF triple contains three components: a *subject*, a *predicate*, and an *object*.

- **Subject**: This is the topic of the RDF statement; it is the resource being described. Resources have properties (i.e., attributes) and relationships. In an RDF statement, an *Internationalized Resource Identifier* (IRI)—a unique character string that helps to identify an entity—is used to unambiguously identify the subject.[76]
- **Predicate**: Properties and relationships are represented by the predicate. The predicate is an indication of the connection between the subject and the object, or it denotes the type of attribute that is expressed in the object. Predicates, too, are identified by IRIs.
- **Object**: This is *another* resource that is connected in some way to the subject, or, it is a value associated with the predicate (e.g., a date). The object may be the IRI for another resource, or it may be a string of alpha-numeric characters (referred to as a *literal*) providing the value for the attribute.

A representation of an RDF triple is presented in Figure 5.4.

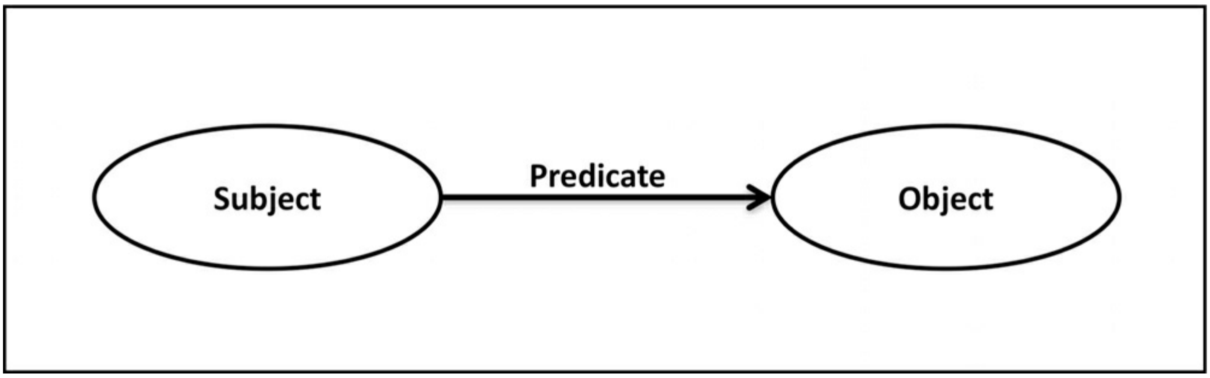


Figure 5.4. A Representation of an RDF Triple.

In Figure 5.5 we have an RDF triple expressing one particular relationship between two resources. The subject of our metadata statement is an online text with the title *A Very Brief Review of the MARC 21 Bibliographic Format,* and it has a creator whose name is known. The unique identifier for the web resource (a URL in this case) represents the subject, the predicate is a creation relationship that might be expressed in any number of ways (here, it is using the Dublin Core element *Creator*), and the object is an identifier that links to the authority record for the creator. IRIs are used for the subject, predicate and object. Although an object can be expressed as a literal, in the RDF model an IRI is preferable whenever possible. The statement in Figure 5.5 is only one piece of a larger description of the document. A full description requires multiple RDF triples to give a complete picture. A set of RDF triples is referred to as an *RDF graph*.

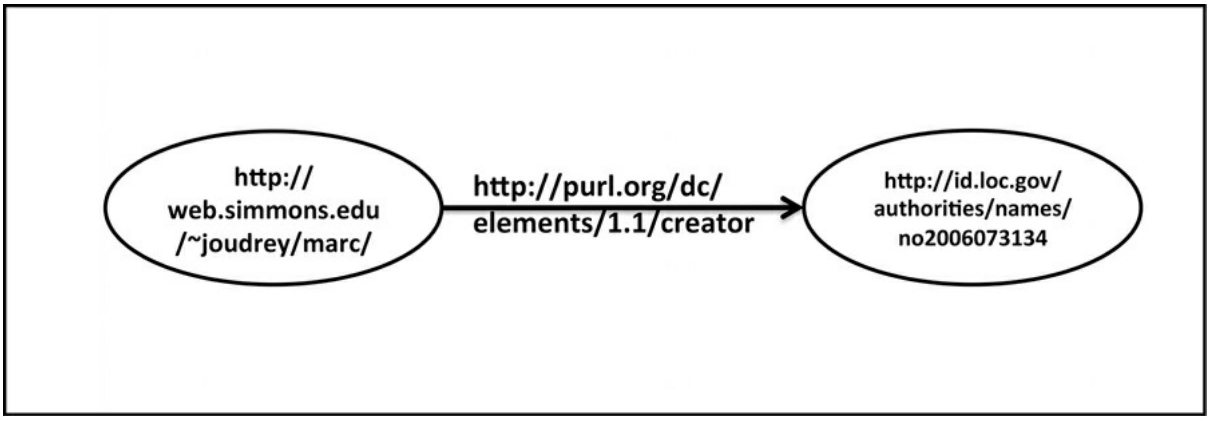


Figure 5.5. An RDF Triple Using URIs.

If the object is another resource, then the second resource could *also* have properties, which would have values, many of which could be *other* resources. There is no limit to the number of connections or links that may exist among the entities. The RDF triple is the basis for what is known as *linked data*—the foundation of the Semantic Web. Figure 5.6 is an illustration of the connections that can exist using linked data based on an RDF structure.

In Figure 5.6 a resource, called *Document 1,* has properties such as *Title*, *Language, Format, Creator,* and *Subject.* Each property has a value. Sometimes, the value is a literal (see the gray rectangles in the figure), and sometimes the value is another entity (see the polka-dotted ovals). The value of the *Creator* attribute is another resource called *Person A.* This resource has properties such as *Name, Position*, *E-mail,* and *Affiliation.* The value of the *Affiliation* property is yet another resource called *Institution X*, which has its own properties (*name* and *location*). The links continue to amass as additional documents are connected to *Person A*—in this case, through creation relationships, but it could be through any number of other relationships as well. And those documents might link to publishers, manufacturers, distributors, and other creators and contributors. As you might imagine, a publisher entity could link to many thousands of other documents, as could subject entities. The notion of the Semantic Web depends on massive amounts of data being linked.

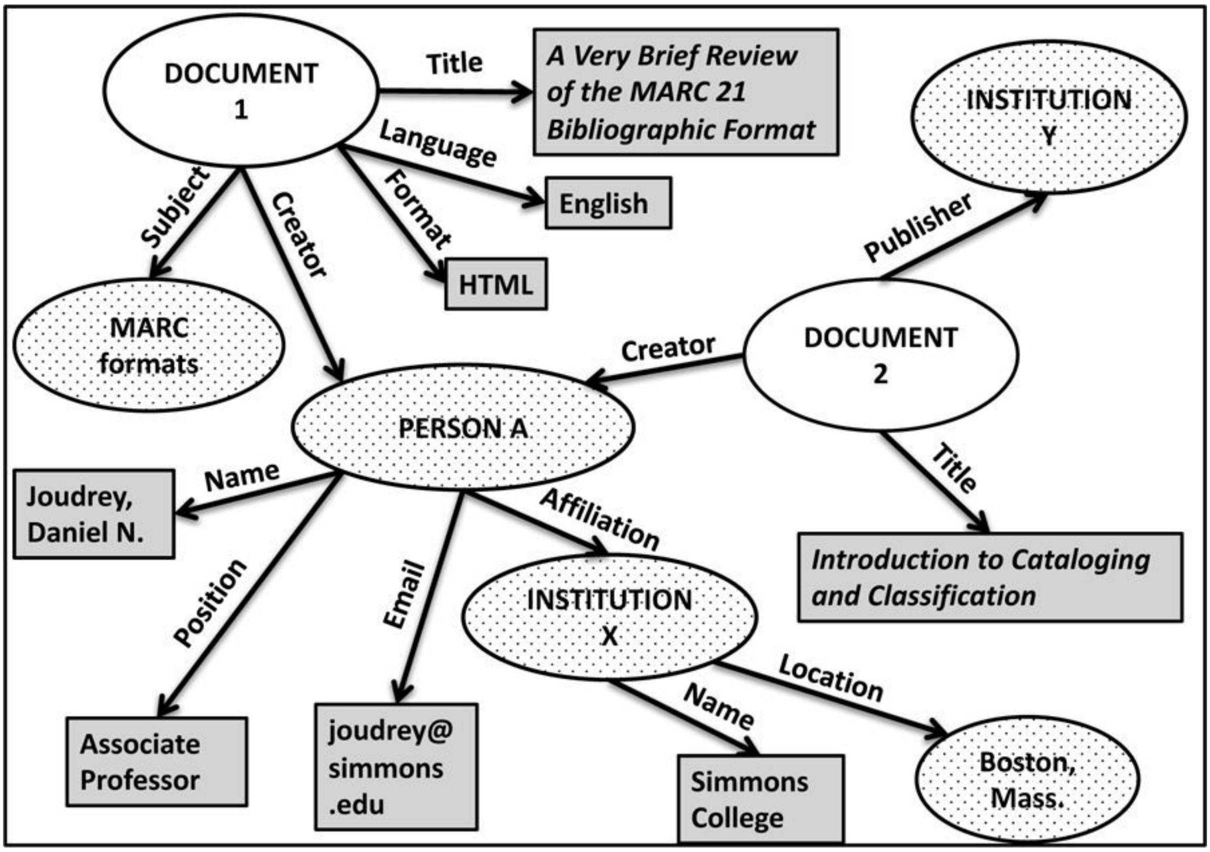


Figure 5.6. Entities Linked with RDF Triples.

In order for this model to be implemented practically, however, it must be expressed concretely. RDF may be encoded in XML[77] or in some other markup language such as Notation3[78] or the Terse RDF Triple Language (Turtle).[79] Using XML, an RDF description identifies the fact that it is using RDF, identifies the namespaces to be used for the property names, and provides the description.

As mentioned above, the statement in Figure 5.5 contains only one RDF triple. This does not reflect the true nature of metadata description. Using RDF, a more complete description of this resource could look like the set of triples found in Table 5.7. Please note that even though this description is more complete, it is still a rather small description of the resource. Generally, more statements would be needed to describe the resource robustly.

Or, the description of the resource could be encoded in XML as seen below. In this view, the descriptive metadata content is highlighted with boldface type, while the encoding and structural elements are in plain type.

Table 5.7. An RDF Graph Describing the Same Resource in Tabular Form.

SUBJECT	PREDICATE	OBJECT
http://web.simmons.edu/~joudrey/marc/	http://purl.org/dc/elements/1.1/title	"A Very Brief Review of the MARC 21 Bibliographic Format"
http://web.simmons.edu/~joudrey/marc/	http://purl.org/dc/elements/1.1/creator	http://id.loc.gov/authorities/names/no20060731
http://web.simmons.edu/~joudrey/marc/	http://purl.org/dc/elements/1.1/language	"en"
http://web.simmons.edu/~joudrey/marc/	http://purl.org/dc/elements/1.1/format	"text/html"
http://web.simmons.edu/~joudrey/marc/	http://purl.org/dc/elements/1.1/subject	http://id.loc.gov/authorities/subjects/sh8508096

```
<?xml version="1.0"?>
<rdf:RDF
        xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
        xmlns:schema="http://schema.org/"
        xmlns:dc="http://purl.org/dc/elements/1.1/"
        xmlns:skos="http://www.w3.org/2004/02/skos/">
<rdf:Description rdf:about="http://web.simmons.edu/~joudrey/marc/
        index.html">
<rdf:type rdf:resource="http://www.w3.org/2001/XMLSchema#
        CreativeWork"></rdf:type>
        <dc:title>A Very Brief Review of the MARC 21
                Bibliographic Format</dc:title>
        <dc:creator rdf:resource="http://id.loc.gov/authorities/
                names/no2006073134"/>
        <dc:language>en</dc:language>
        <dc:format>text/html</dc:format>
        <dc:subject rdf:resource="http://id.loc.gov/authorities/
                subjects/sh85080966"/>
</rdf:Description>

<rdf:Description
rdf:about="http://id.loc.gov/authorities/names/no2006073134">
        <schema:familyName>Joudrey</schema:familyName>
        <schema:givenName>Daniel</schema:givenName>
</rdf:Description>

<rdf:Description
rdf:about="http://id.loc.gov/authorities/subjects/sh85080966">
        <skos:inScheme>LCSH</skos:inScheme>
        <skos:prefLabel>MARC formats</skos:prefLabel>
        <skos:altLabel>Machine-Readable Cataloging
                formats</skos:altLabel>
</rdf:Description>
</rdf:RDF>
```

In this example, the first two lines state that this description is in XML, using RDF as the metadata structure. The next four lines identify that the tags used throughout the description come from RDF itself, but also from the Schema.org vocabulary, the Dublin Core Metadata Element Set, and the Simple Knowledge Organization System (SKOS); IRIs for their namespaces are included. The next line, beginning with "rdf:Description rdf:about," indicates which resource is being described. It provides the IRI for *A Very Brief Review of the MARC 21 Bibliographic Format* and the type of resource that it is. What then follows is the description of the resource using five Dublin Core elements as predicates as was seen in the triples listed in Table 5.7. The first "</rdf:Description>" closing tag designates the end of the metadata describing the website.

Please note that this website is not the only thing being described in the metadata though; two smaller RDF graphs follow. The first enhances the description of the creator, providing a given and a family name to supplement the standard identifier used. The second provides more information about the subject matter of the original resource, showing the preferred term and an alternative term, as well as the source vocabulary that was used. A group of RDF graphs such as this is referred to as an *RDF dataset*. This RDF dataset can be represented graphically, as seen in Figure 5.7.

As stated, RDF is considered one of the major building blocks in the vision for the Semantic Web, an idea first articulated in 1997 by Tim Berners-Lee, the person who is credited with inventing the World Wide Web.[80] The Semantic Web is intended to provide more structure to the web that will allow computers to deal with its content in meaningful ways. It is intended that information will be defined in such a way that its meaning or semantics can be discernible, shared, and processed by automated tools as well as by people. This data, formatted as RDF triples, is stored on the web in *triplestores*, which are databases specifically designed to house RDF structured data for Semantic Web queries.[81] Berners-Lee and colleagues give an example of a brother and sister needing to coordinate their schedules with a certain kind of physical therapy specialist who is also covered by a particular insurance plan so that one of them can take their mother to this specialist. The Semantic Web allows their automated agents to work together to find the right place at a time one of them has available, all within a few minutes.[82] Berners-Lee and colleagues state, "For the semantic web to function, computers must have access to structured collections of information and sets of inference rules that they can use to conduct automated reasoning."[83] The RDF metadata model provides this structure. Several groups within W3C are working toward realization of the Semantic Web, including several working on RDF-related specifications.[84]

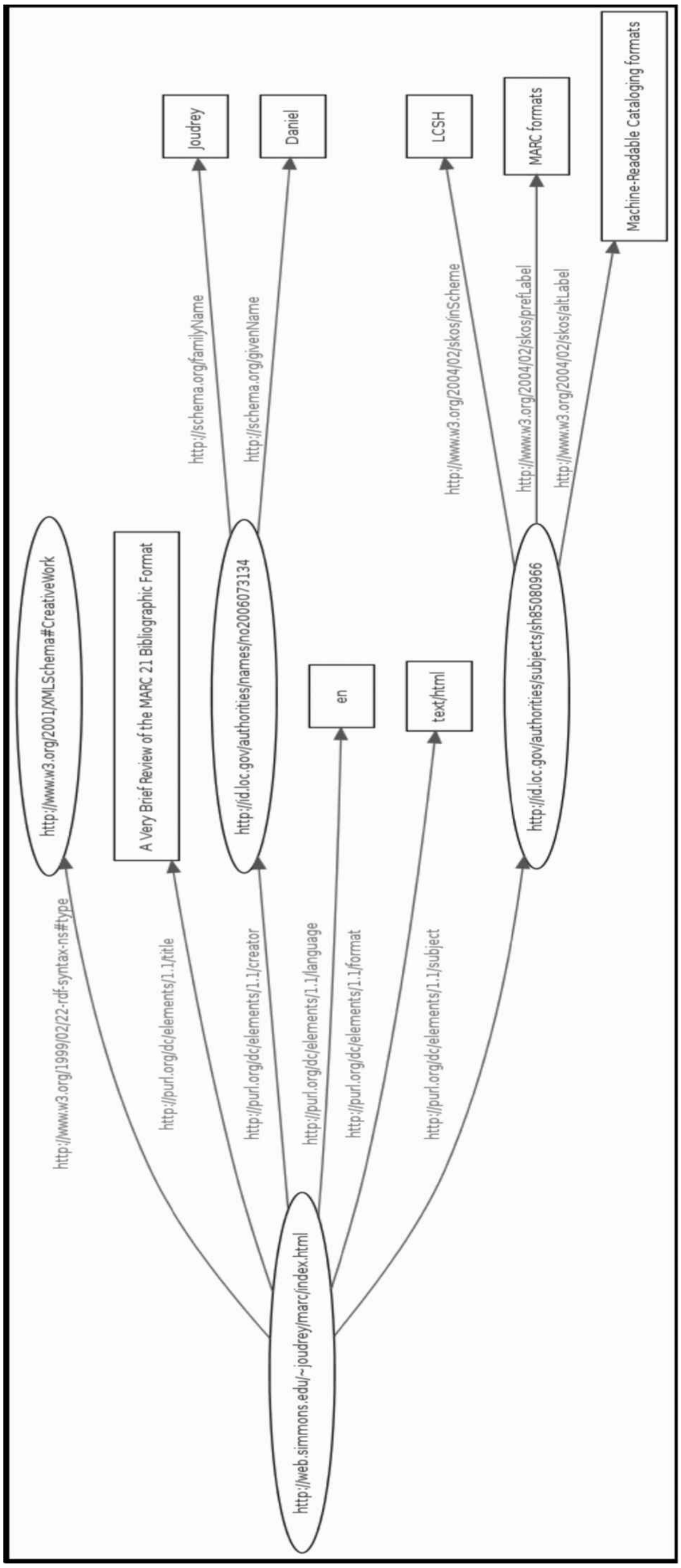


Figure 5.7. An RDF Dataset in Graphic Form.

This section is just a brief overview of some of the basic concepts associated with RDF and its use in the linked data landscape of the Semantic Web. There is much more to RDF. For more complete information about the Semantic Web, linked data, and the RDF family of standards, please consult the ample, but dense, documentation provided by the W3C on the web.[85]

DCMI Abstract Model (DCAM)

RDF has influenced the DCMI Abstract Model, published in 2007 by the Dublin Core Metadata Initiative (DCMI). The primary purpose of the DCAM is to provide a conceptual model of metadata that is independent of any specific approach to encoding or semantics. By specifying the components and offering a generic model, the DCAM allows for better communications about metadata and interoperability across information environments. It describes how the syntax-independent model's entities are related in order to create complete metadata structures. The model "is primarily aimed at the developers of software applications that support Dublin Core metadata, people involved in developing new syntax encoding guidelines for Dublin Core metadata and people developing metadata application profiles based on DCMI vocabularies or on other compatible vocabularies."[86]

The abstract model comprises three related information models: a *resource model*, a *description set model*, and a *vocabulary model*. The three information models contain entities of various types. The DCMI Resource Model includes entities such as *resources, properties,* and *values* (which should seem familiar after reading about RDF and entity-relationship models). As is the case with RDF, a resource can be any *thing* in the universe: a document, a person, an abstract concept, and so on. Resources are described using the same basic model, which is summarized entirely by a few simple statements in the DCAM.

- Each *described resource* is described using one or more *property-value pairs.*
- Each *property-value pair* is made up of one *property* and one *value.*
- Each *value* is a *resource*—the physical, digital or conceptual entity or *literal* that is associated with a *property* when a *property-value pair* is used to describe a *resource.* Therefore, each *value* is either a *literal value* or a *non-literal value:*
    - A *literal value* is a *value* which is a *literal.*
    - A *non-literal value* is a *value* which is a physical, digital or conceptual entity.
- A *literal* is an entity which uses a Unicode string as a lexical form, together with an optional language tag or datatype, to denote a resource.[87]

The DCMI Resource Model states that Dublin Core metadata consists of *statements* describing resources and each description refers to properties and values (a second resource). This is analogous to the resource-property-value model (i.e., the subject–predicate–object structure) found in RDF triples. The concept of the *described resource* is equivalent to the *subject* in the RDF model. The importance of the property–value (i.e., predicate–object) relationship cannot be overstated. It is not enough to know either the property *or* the value in isolation; it is the combination that makes that metadata useful. For example, it is not enough to know that the resource is associated with Tokyo in some fashion. It is important to understand that Tokyo is the place of distribution or that it is the topic of the resource. The meaning of that geographic name is very different depending on the context. The context is provided by identifying the property (i.e., attribute, element, or relationship) unambiguously, and associating it clearly with the value "Tokyo."

The DCMI Description Set Model introduces additional entities: *descriptions* and *description sets.* So, if a metadata statement is analogous to the *RDF triple*, then a *description* is analogous to an *RDF graph* for a single resource. This is based on the Dublin Core *One-to-One Principle*, which states that a description should be focused on one (and only one) resource at a time. For example, the metadata for *A Very Brief Review of the MARC 21 Bibliographic Format* outlined in Table 5.7 is a *description* in DCAM. A *description set*, then, is two or more metadata descriptions (similar to an *RDF dataset*). In the DCAM, a *record* contains one or more description sets.

The final component is the DCMI Vocabulary Model, which introduces the concepts: *vocabularies, terms, properties, classes, vocabulary encoding schemes,* and *syntax encoding schemes.* In short, *terms* make up *vocabularies*. Terms may be in the form of *properties*, *classes*, *syntax encoding schemes*, or *vocabulary encoding schemes*. For more about the vocabulary model or any other part of the DCAM, please consult the latest version on the Dublin Core website.

The future of the DCAM is not particularly clear. It has seen little further development since it was originally presented, and its reception has not been particularly enthusiastic. In a blog post in 2008, library technologist Karen Coyle expresses this sentiment clearly:

> Few outside of the small group that worked on the DCAM profess to understand it. I myself have read and re-read the DCAM document dozens of times and have still failed to have the "aha!" moment in which I would see how the DCAM explains everything of importance to metadata creation.[88]

Part of the problem may be that much of the ground it covers is already found in the more familiar and more widely implemented RDF model. In fact, some discussions within the DCMI in 2009–2010 proposed, among other things, deprecating the DCAM in favor of RDF. Others proposed removing two of the three components of the abstract model (leaving only the Description Set Model) again in favor of RDF.[89]

METADATA AND CATALOGING

In recent years some have debated whether *metadata* was a new concept in the online age or whether it was simply a recasting of activities performed in libraries for centuries. Some discussions of metadata refer exclusively to electronic materials, but the term is not necessarily restricted to digital resources only. Many authors like to point out that the library profession has been creating metadata for millennia. Even the earliest Sumerian lists contained metadata in some form. Some authors equate creating metadata for electronic resources with the cataloging of books. Some consider cataloging to be a subset of activities under the broader concept of metadata creation.

Parallels between the two processes certainly exist. The basic objectives of metadata creation and cataloging (i.e., providing descriptions of and access to items) are alike. The processes used to create the descriptions are also similar; both focus on recording the attributes that allow users to identify and select the information resources that most closely meet their needs. In addition, electronic and analog materials share many characteristics. Both generally have titles, creators, creation dates, subject matter, and publication sources of some sort. There are enough similarities between the two activities and the information resources treated by them to see that a relationship exists between metadata creation and cataloging.

While there are parallels, differences are also claimed. Some feel the dissimilarities are too great to equate metadata with cataloging. Stefan Gradmann states that the differences lie in

- who creates the metadata (nonprofessionals),
- why it is created (resource discovery, not just description),
- the process (more efficiently produced), and
- the materials described (electronic resources).[90]

It is questionable that these are valid distinctions. Just as the ideal of having authors create their own metadata has not come to pass, it is not just professional catalogers who are generating cataloging data. For decades, *paraprofessionals* (library staff members without a master's degree in library and information science) have done much cataloging. Cataloging, too, is about resource discovery. That is why there has been such an emphasis on access points for names and subjects with a concentration on authority control for these. Classification provides yet another way for users to discover resources by finding them alongside other materials on the same subject. With respect to Gradmann's third difference, *the process,* cataloging became more efficient in recent decades with its greater use of information technology and with the transfer of the major cataloging tools to the web platform. Finally, the conventions of cataloging have always been expanded to apply to information resources in new formats (digital or otherwise), as seen in the latest version of the cataloging guidelines found in RDA, which strives to be format neutral.[91] In many ways, it appears that metadata creation is just an extension of traditional cataloging.

More useful distinctions that have been raised tend to focus on certain characteristics of online resources. These differences are important to acknowledge, even though they do not negate the relationship between cataloging and metadata creation. The first of these is more perception than reality: it is the difficulty of determining what is an information resource. With the explosion of new types of resources, it introduces a difficulty in deciding what is an "analyzable or describable unit." The cataloging world, however, has had centuries to develop standards and practices. For the most part, one physical item has been equated with one bibliographic description. Although there are questions about when to provide a single record for multipart resources (e.g., serials, multivolume sets) and when to provide multiple records for just one resource (e.g., anthologies of short stories, articles in journals), these questions are not new. In the twenty-first century, we find the same problems for digital resources. What is a describable unit in a digital library? Do we create metadata for a single page or for the entire website that contains it? Do we describe a digital image or a collection of digital images? This returns us to the granularity issues mentioned earlier. In addition, with a digital image, what exactly is being described? Should the metadata describe the digital image only? Should it describe both the digitized photo and the original physical photograph? If both are in the collection, should there be one record or two?

An obvious difference between tangible and digital resources is that remote-access online resources have no physical carriers, unlike books, maps, or a set of Blu-Ray discs. A key feature of traditional cataloging is the physical description of the item in hand (e.g., 280 pages : illustrations ; 26 cm *or* 1 videocassette (15 min.) : sound, color ; 3/4 in.). Except, perhaps, for indications of illustrative matter, color, and sound, a physical description of a blog seems fairly unnecessary.

Another distinction between digital and tangible resources relates to the concept of edition. For some electronic resources, like commercial software, it is obvious when a new edition or version has been released. But for others, such as websites, it is not so obvious. Online content can be ephemeral and/or unstable. Some resources simply disappear without a trace or may be hidden from sight (although many of these can be found in the Internet Archive). Others can, and do, change often. While some sites may have relatively stable content, others are updated frequently.

Location information also differs. When metadata records and information resources are both encoded and accessible online, the encoded nature of the metadata records allows users to locate and access an information resource almost simultaneously. Therefore, special attention must be paid to location/access metadata for electronic resources. URLs are far more likely to change than are call numbers in an average library, and keeping them up to date may be time consuming and difficult.

The last major difference between tangible and digital resources that we will discuss is structural. Compared to the structure of a physical resource, the structure of a digital object can be very complex. For example, a digitized book requires more and different types of metadata than does a physical book sitting on a shelf. In order for it to be found and used, the physical book, for the most part, requires only descriptive metadata (such as title, author, and subject) and some kind of location device (e.g., a call number), with some administrative metadata thrown in for good measure. The digital object, however, requires extensive structural metadata in order for it to be displayed and to function properly in an online environment. This is *in addition* to the necessary descriptive metadata about the digitized resource, the original source, and any required administrative metadata (such as technical, preservation, rights, and meta-metadata). Of the differences between digital and tangible resources, this is the one that has the greatest impact on activities related to organizing information resources. It is the strongest argument in favor of distinguishing between metadata creation and cataloging, although it is only applicable to a portion of electronic resources.

CONCLUSION

This chapter has provided a simple introduction to metadata issues. Although many see metadata creation as little more than the activities of traditional library cataloging applied to new formats, there are some additional considerations required for digital resources. The complexity, quantity, and variety of information required to organize digital resources is often greater than those for traditional resources, and at times, new approaches are needed. It is important to remember, though, that the ultimate purpose for both cataloging and metadata creation is to allow users to navigate information systems to find, identify, select, obtain, and explore the information they need. In addition, it is important that we are careful in the terminology we use when discussing metadata. Although the conceptual components discussed throughout this chapter are highly intertwined, it can be helpful to specify the type of information we are referring to when we use the term *metadata.* One person's metadata schema may be another person's structure standard.

Some Important Terms in This Chapter

Administrative Metadata
Agency
Agent
Application Profile
Collective Agent
Contextualize
Controlled Access Point
Controlled Vocabulary
Crosswalk
Data Value Standard
Description
Descriptive Metadata
Element
Entity
Entity–Relationship Model
Explore
Expression
Extensibility
Find
Flexibility
Granularity
Identifier
Identify
Interoperability
Item
Justify
Linked Data
Literal
Manifestation
Meta Tag
Metadata
Metadata Registry
Meta-metadata
Model
Name
Namespace
Navigate
Nomen
Non-literal
Obtain
Preservation Metadata
Property
Property–Value Pair
Qualifier
RDF Dataset
RDF Graph
Record
Res
Resource
Rights and Access Metadata
Rules
Schema
Select
Semantic Web
Semantics
Structural Metadata
Structure
Syntax
Technical Metadata
Thema
Time-Span
Triple
Triplestore
User Tasks
Value
Work

Some Important Acronyms in This Chapter

CIDOC:	International Committee on Documentation
CRM:	Conceptual Reference Model
DACS:	*Describing Archives: A Content Standard*
DC:	Dublin Core
DCAM:	DCMI Abstract Model
DCMI:	Dublin Core Metadata Initiative
DDI:	Data Documentation Initiative
EAD:	Encoded Archival Description
EGAD:	Experts Group on Archival Description
Exif:	Exchangeable Image File Format
FRAD:	*Functional Requirements for Authority Data*
FRBR:	*Functional Requirements for Bibliographic Records*
FRSAD:	*Functional Requirements for Subject Authority Data*
IFLA:	International Federation of Library Associations and Institutions
IFLA LRM:	*IFLA Library Reference Model*
IRI:	Internationalized Resource Identifier
ISBD:	*International Standard Bibliographic Description*
LRM:	*Library Reference Model*
METS:	Metadata Encoding & Transmission Standard
MIX:	NISO Metadata for Images in XML Schema
MODS:	Metadata Object Description Schema
PREMIS:	Preservation Metadata: Implementation Strategies
RDA:	*Resource Description & Access*
RDF:	Resource Description Framework
RiC:	Records in Context
TEI:	Text Encoding Initiative
VRA:	Visual Resource Association
W3C:	World Wide Web Consortium
XML:	Extensible Markup Language

NOTES

All URLs accessed June 2017.

1. "Metadata," *FOLDOC: Free On-line Dictionary of Computing*, edited by Denis Howe, http://foldoc.org/metadata.

2. Information resources and metadata records are addressed throughout this book, but identifying attributes are primarily addressed in Chapter 7, as are metadata elements, structure standards, and content standards. Value standards that are not subject-based are addressed in Chapter 8, and subject-related value standards are addressed in Chapters 10 and 11. Encoding is addressed in Chapter 6 and data models are addressed later in *this* chapter.

3. These categories of metadata are described in detail later in this chapter and in Chapter 7.

4. Anne J. Gilliland, "Setting the Stage," in *Introduction to Metadata*, 3rd ed., edited by Murtha Baca (Los Angeles: Getty Research Institute, 2016), http://www.getty.edu/publications/intrometadata/setting-the-stage/.

5. U.S. National Archives and Records Administration (NARA), "Technical Guidelines for Digitizing Archival Materials for Electronic Access: Creation of Production Master Files–Raster Images," June 2004, http://www.archives.gov/preservation/technical/guidelines.pdf.

6. Katherine M. Wisser, "Introduction to the Special Issue: Contextual Metadata," *Journal of Library Metadata* 11, no. 3–4 (2011): 101–3; Joan E. Beaudoin, "A Framework for Contextual Metadata Used in the Digital Preservation of Cultural Objects," *D-Lib Magazine* 18, no. 11–12 (2012), http://www.dlib.org/dlib/november12/beaudoin/11beaudoin2.html.

7. Lynne C. Howarth, "Metadata and Bibliographic Control: Soul-Mates or Two Solitudes?" *Cataloging & Classification Quarterly* 40, no. 3–4 (2005): 37–56.

8. Sherry L. Vellucci, "Metadata and Authority Control," *Library Resources & Technical Services* 44, no. 1 (2000): 33–43.

9. Stuart Weibel, "Metadata: Semantics; Structure; Syntax," in "Weibel Lines: Ruminations on Libraries and Internet Standards," http://weibel-lines.typepad.com/weibelines/2008/02/metadata-semant.html.

10. Ibid.

11. Vellucci, "Metadata and Authority Control," 36–37.

12. Ibid., 36.

13. "PREMIS: Preservation Metadata Maintenance Activity," Library of Congress, http://www.loc.gov/standards/premis/.

14. "MIX: NISO Metadata for Images in XML Schema: Technical Metadata for Digital Still Images Standard," Library of Congress, http://www.loc.gov/standards/mix/.

15. *PREMIS Data Dictionary for Preservation Metadata,* Version 3.0, Library of Congress, 32, http://www.loc.gov/standards/premis/v3/premis-3-0-final.pdf.

16. For the latest developments on the ODRL, check out the ODRL Community Group page, https://www.w3.org/community/odrl/.

17. Jytte Hansen and Leif Andresen, "AC-Administrative Components: Dublin Core DCMI Administrative Metadata," Dublin Core Metadata Initiative, October 28, 2003, http://dublincore.org/groups/admin/ or http://www.bs.dk/standards/AdministrativeComponents.htm.

18. "Digital Page Imaging and SGML: An Introduction to the Electronic Binding DTD (Ebind)," Berkeley Digital Library SunSITE, http://sunsite.berkeley.edu/Ebind/.

19. Although the old MOA2 site is down, it is discussed at "Technology Reports: The Making of America II Project," *Cover Pages,* January 7, 2002, http://xml.coverpages.org/moa2.html; "The Making of America, Part 2," Digital Library Federation, 2004, https://old.diglib.org/standards/dlfmoaii.htm.

20. "METS: Metadata Encoding & Transmission Standard," Library of Congress, http://www.loc.gov/standards/mets/.

21. "METS Example Documents," Library of Congress, http://www.loc.gov/standards/mets/mets-examples.html.

22. Priscilla Caplan, *Metadata Fundamentals for All Librarians* (Chicago: American Library Association, 2003), 7.

23. DCMI-Libraries Working Group, "Library Application Profile," Dublin Core Metadata Initiative, http://dublincore.org/documents/library-application-profile/.

24. For example, see the "Dublin Core Collections Application Profile," http://dublincore.org/groups/collections/collection-application-profile/ or the "Library Application Profile" listed in the previous endnote.

25. International Organization for Standardization, "ISO/IEC 11179-3:2013(E): Information Technology–Metadata Registries (MDR)–Part 3: Registry Metamodel and Basic Attributes," http://metadata-standards.org/11179/.

26. "Open Metadata Registry," http://metadataregistry.org/.

27. Margaret St. Pierre and William P. LaPlant, "Issues in Crosswalking Content Metadata Standards," NISO, http://www.niso.org/publications/white_papers/crosswalk/.

28. Library of Congress, Network Development and MARC Standards Office, "Dublin Core Metadata Element Set Mapping to MODS Version 3," http://www.loc.gov/standards/mods/dcsimple-mods.html.

29. Open Archives Initiative, "Open Archives Initiative Protocol for Metadata Harvesting," http://www.openarchives.org/pmh/.

30. "Model," *FOLDOC: Free On-line Dictionary of Computing*, edited by Denis Howe, http://foldoc.org/model.

31. "Model," Merriam-Webster.com, http://www.merriam-webster.com/dictionary/model.

32. International Federation of Library Associations and Institutions, IFLA Study Group, *Functional Requirements for Bibliographic Records: Final Report* (Munich: Saur, 1998), http://www.ifla.org/publications/functional-requirements-for-bibliographic-records.

33. Entity-relationship models were first proposed by Peter Chen in 1976 in the context of database design. For more information see, Peter Pin-Shan Chen, "The Entity-Relationship Model: Toward a Unified View of Data," *ACM Transactions on Database Systems* 1, no. 1 (1976): 9–36.

34. IFLA Working Group on Functional Requirements and Numbering of Authority Records (FRANAR), *Functional Requirements for Authority Data* [FRAD]: *A Conceptual Model*, approved by the Standing Committees of the IFLA Cataloguing Section and IFLA Classification and Indexing Section, 2009, amended and corrected through July 2013, http://www.ifla.org/publications/functional-requirements-for-authority-data.

35. IFLA Working Group on the Functional Requirements for Subject Authority Records (FRSAR), *Functional Requirements for Subject Authority Data (FRSAD): A Conceptual Model*, approved by the Standing Committee of the IFLA Section on Classification and Indexing, 2010, http://www.ifla.org/node/5849.

36. Pat Riva, Patrick Le Boeuf, and Maja Žumer, "IFLA Library Reference Model: A Conceptual Model for Bibliographic Information," August 2017 (The Hague: IFLA, 2017), https://www.ifla.org/publications/node/11412.

37. Daniel N. Joudrey, Arlene G. Taylor, and David P. Miller, *Introduction to Cataloging and Classification*, 11th ed. (Santa Barbara, CA: Libraries Unlimited, 2015), 51–124.

38. Charles A. Cutter, *Rules for a Dictionary Catalog*, 4th ed. (Washington, DC: Government Printing Office, 1904; reprint, London: The Library Association, 1962), 12.

39. International Conference on Cataloguing Principles, Paris, 9th–18th October, 1961), *Report* (London: International Federation of Library Associations, 1963), 91–96. Also available in A. H. Chaplin, "IFLA International Conference on Cataloguing Principles. Preliminary Official Report," *Library Resources & Technical Services* 6, no. 2 (1962): 161–70; *Statement of Principles Adopted at the International Conference on Cataloguing Principles, Paris, October, 1961,* annotated ed., with commentary and examples by Eva Verona (London: IFLA Committee on Cataloguing, 1971); and A. H. Chaplin and Dorothy Anderson, eds. *Report/International Conference on Cataloguing Principles, Paris, 9th–18th October 1961* (London: IFLA International Office for UBC, 1981).

40. FRBR, 79, http://www.ifla.org/files/assets/cataloguing/frbr/frbr_2008.pdf.

41. Elaine Svenonius, *The Intellectual Foundation of Information Organization* (Cambridge, MA: MIT Press, 2000), 18–19.

42. FRAD, 46, http://www.ifla.org/files/assets/cataloguing/frad/frad_2013.pdf.

43. FRSAD, 33–34, http://www.ifla.org/files/assets/classification-and-indexing/functional-requirements-for-subject-authority-data/frsad-final-report.pdf.

44. Ibid.

45. International Federation of Library Associations and Institutions (IFLA), Meetings of Experts on an International Cataloguing Code, *Statement of International Cataloguing Principles*, 2016 edition (The Hague:

IFLA, 2016), 10, http://www.ifla.org/files/assets/cataloguing/icp/icp_2016-en.pdf.

46. FRBR, 13, http://www.ifla.org/files/assets/cataloguing/frbr/frbr_2008.pdf.

47. This figure was first used in Joudrey, Taylor, and Miller's *Introduction to Cataloging and Classification*, 66.

48. FRBR, 19, http://www.ifla.org/files/assets/cataloguing/frbr/frbr_2008.pdf.

49. Ibid.

50. Barbara B. Tillett, "Bibliographic Relationships," in *Relationships in the Organization of Knowledge,* eds. Carol A. Bean and Rebecca Green (Dordrecht, The Netherlands: Kluwer Academic, 2001), 19–35.

51. Richard P. Smiraglia, "Authority Control and the Extent of Derivative Bibliographic Relationships" (PhD diss., University of Chicago, 1992).

52. FRBR, 55–78, http://www.ifla.org/files/assets/cataloguing/frbr/frbr_2008.pdf.

53. IFLA, *Statement of International Cataloguing Principles*, 2016 edition, http://www.ifla.org/files/assets/cataloguing/icp/icp_2016-en.pdf.

54. *RDA: Resource Description & Access*, available as a subscription product through the RDA Toolkit, http://www.rdatoolkit.org/.

55. Riva, Le Boeuf, and Žumer, "IFLA Library Reference Model," 5, https://www.ifla.org/publications/node/11412.

56. Ibid., 15.

57. Ibid., 15–16.

58. Pat Riva, Patrick Le Boeuf, and Maja Žumer, "Transition Mappings: User Tasks, Entities, Attributes, and Relationships in FRBR, FRAD, and FRSAD Mapped to Their Equivalents in the IFLA Library Reference Model," August 2017, 2, https://www.ifla.org/publications/node/11412.

59. Ibid.

60. Riva, Le Boeuf, and Žumer, "IFLA Library Reference Model," 37, https://www.ifla.org/publications/node/11412.

61. Ibid., 61–82.

62. Murtha Baca and Sherman Clarke, "FRBR and Works of Art, Architecture, and Material Culture," in *Understanding FRBR: What It Is, and How It Will Affect Our Retrieval Tools,* ed. Arlene G. Taylor (Westport, CT: Libraries Unlimited, 2007), 103–10.

63. Gretchen Gueguen et al., "Toward an International Conceptual Model for Archival Description: A Preliminary Report from the International Council on Archives' Expert Group on Archival Description," *American Archivist* 76, no. 2 (2013): 567–84.

64. International Council of Museums, International Committee for Documentation [ICM CIDOC], "New Version of ISO 211127 is Announced," in the *CIDOC-Conceptual Reference Model* website, http://cidoc-crm.org/Event/new-version-of-iso-211127-is-announced.

65. ICM CIDOC, "What Is the CIDOC CRM?" in the *CIDOC-Conceptual Reference Model* website, http://cidoc-crm.org/.

66. ICM CIDOC, "Definition of the CIDOC Conceptual Reference Model," in the *CIDOC-Conceptual Reference Model* website, vii–ix, http://cidoc-crm.org/Version/version-6.2.2.

67. ICM CIDOC, "Functional Overview," in the *CIDOC-Conceptual Reference Model* website, http://cidoc-crm.org/functional-units.

68. These and other mappings can be found at: ICM CIDOC, "CIDOC CRM Mappings," in the *CIDOC-Conceptual Reference Model* website, http://www.cidoc-crm.org/node/209.

69. ICM CIDOC, "The CIDOC CRM Home Page," http://cidoc-crm.org/.

70. For more discussion of the Warwick Framework, see Susan S. Lazinger, *Digital Preservation and Metadata: History, Theory, Practice* (Englewood, CO: Libraries Unlimited, 2001), 193–95; Carl Lagoze, "The Warwick Framework: A Container Architecture for Diverse Sets of Metadata," *D-Lib Magazine* 2 (July/August 1996), http://www.dlib.org/dlib/july96/lagoze/07lagoze.html.

71. In this section, we are primarily concerned with the conceptual model rather than full implementations of RDF. Since the original version of RDF was published, the framework has been expanded considerably to include specific recommendations on syntax and vocabulary. Describing the full range of RDF activities and applications is beyond the scope of this book.

72. W3C, "RDF," February 25, 2014, https://www.w3.org/RDF/.

73. W3C, "RDF 1.1 Primer," February 25, 2014, https://www.w3.org/TR/2014/NOTE-rdf11-primer-20140225/.

74. W3C, "RDF 1.1 Concepts and Abstract Syntax," February 25, 2014, https://www.w3.org/TR/rdf11-concepts/.

75. W3C, "RDF 1.1 Primer," February 25, 2014, https://www.w3.org/TR/2014/NOTE-rdf11-primer-20140225/.

76. An IRI is similar to a Uniform Resource Identifier (URI), but works with Unicode characters, not just ASCII characters. It is an improvement to the URI.

77. W3C, "RDF 1.1 XML Syntax," February 25, 2014, http://www.w3.org/TR/rdf-syntax-grammar/.

78. W3C, "Notation 3 (N3): A Readable RDF Syntax," March 28, 2011, https://www.w3.org/TeamSubmission/n3/.

79. W3C, "RDF 1.1 Turtle—Terse RDF Triple Language," February 25, 2014, https://www.w3.org/TR/turtle/.

80. Tim Berners-Lee, "Realising the Full Potential of the Web," http://www.w3.org/1998/02/Potential.html; and Tim Berners-Lee, James Hendler, and Ora Lassila, "The Semantic Web," *Scientific American* 284, no. 5 (May 2001): 34–38, 40–43.

81. Jack Rusher, "Triple Store," SWAD-Europe Workshop on Semantic Web Storage and Retrieval Position Papers, 13–14 November 2003, Vrije Universiteit, Amsterdam, Netherlands, https://www.w3.org/2001/sw/Europe/events/20031113-storage/positions/rusher.html.

82. Berners-Lee, Hendler, and Lassila, "The Semantic Web."

83. Ibid.

84. W3C, "W3C Data Activity: Building the Web of Data," https://www.w3.org/2013/data/.

85. W3C, "RDF Current Status," https://www.w3.org/standards/techs/rdf#w3c_all.

86. Andy Powell, Mikael Nilsson, Ambjörn Naeve, Pete Johnston, and Thomas Baker, "DCMI Abstract Model," Dublin Core Metadata Initiative, issued June 4, 2007, http://dublincore.org/documents/abstract-model/.

87. Ibid.

88. Karen Coyle, "DCAM Explained . . . or, An Attempt to Unravel Key Concepts," *Karen Coyle's Home Page*, 2008, http://kcoyle.net/dcam.html.

89. Thomas Baker and Pete Johnston, "A Review of the DCMI Abstract Model with Scenarios for Its Future," written October 15, 2010, revised May 12, 2011, http://wiki.dublincore.org/index.php/Review_of_DCMI_Abstract_Model.

90. Stefan Gradmann, "Cataloging vs. Metadata: Old Wine in New Bottles?" *International Cataloguing and Bibliographic Control* 28, no. 4 (1999): 88–90.

91. *RDA: Resource Description & Access*, available as a subscription product through the RDA Toolkit, http://www.rdatoolkit.org/.

SUGGESTED READINGS

Baca, Murtha, ed. *Introduction to Metadata.* 3rd ed. Los Angeles: Getty Research Institute, 2016. http://www.getty.edu/publications/intrometadata/.

Caplan, Priscilla. *Metadata Fundamentals for All Librarians.* Chicago: American Library Association, 2003. Chapter 16: "Administrative Metadata," Chapter 17: "Structural Metadata," and Chapter 18: "Rights Metadata."

Hillmann, Diane, and Elaine L. Westbrooks, eds. *Metadata in Practice.* Chicago: American Library Association, 2004.

International Federation of Library Associations and Institutions, IFLA Study Group. *Functional Requirements for Bibliographic Records: Final Report.* Munich: Saur, 1998. http://www.ifla.org/publications/functional-requirements-for-bibliographic-records.

International Federation of Library Associations and Institutions, Working Group on FRANAR. *Functional Requirements for Authority Data: A Conceptual Model.* Approved by the Standing Committees of the IFLA Cataloguing Section and IFLA Classification and Indexing Section, 2009, amended and corrected through July 2013. http://www.ifla.org/publications/functional-requirements-for-authority-data.

International Federation of Library Associations and Institutions, Working Group on the Functional Requirements for Subject Authority Records (FRSAR). *Functional Requirements for Subject Authority Data: A Conceptual Model.* Approved by the Standing Committee of the IFLA Section on Classification and Indexing, 2010. http://www.ifla.org/node/5849.

Intner, Sheila S., Susan S. Lazinger, and Jean Weihs. *Metadata and Its Impact on Libraries.* Westport, CT: Libraries Unlimited, 2006. Chapter 1: "What Is Metadata?" and Chapter 2: "Metadata Schemas and Their Relationships to Particular Communities."

Jin, Qiang. *Demystifying FRAD: Functional Requirements for Authority Data.* Santa Barbara, CA: Libraries Unlimited, 2012.

Joudrey, Daniel N., Arlene G. Taylor, and David P. Miller. *Introduction to Cataloging and Classification.* 11th ed. Santa Barbara, CA: Libraries Unlimited, 2015. Chapter 3: "Description and Access: Underlying Principles and Conceptual Models."

Maxwell, Robert L. *FRBR: A Guide for the Perplexed.* Chicago: American Library Association, 2008.

Mitchell, Erik T. *Library Linked Data: Early Activity and Development.* Chicago: ALA TechSource, 2016.

Mitchell, Erik T. *Metadata Standards and Web Services in Libraries, Archives, and Museums.* Santa Barbara: Libraries Unlimited, 2015.

Pomerantz, Jeffrey. *Metadata.* Cambridge, MA: MIT Press, 2015.

Riley, Jenn. *Understanding Metadata.* Bethesda, MD: National Information Standards Organization, 2017. http://www.niso.org/apps/group_public/download.php/17446/Understanding%20Metadata.pdf.

Riva, Pat, Patrick Le Boeuf, and Maja Žumer. "IFLA Library Reference Model: A Conceptual Model for Bibliographic Information." The Hague, Netherlands: IFLA, August 2017. https://www.ifla.org/publications/node/11412.

Smiraglia, Richard P., ed. *Metadata: A Cataloger's Primer.* New York: Haworth Information Press, 2005.

Taylor, Arlene G., ed. *Understanding FRBR: What It Is and How It Will Affect Our Retrieval Tools.* Westport, CT: Libraries Unlimited, 2007.

Zeng, Marcia Lei, and Jian Qin. *Metadata.* 2nd ed. Chicago: ALA Neal-Schuman, 2016.

CHAPTER 6

ENCODING STANDARDS

In the digital information environment in which we live in the twenty-first century, metadata must be encoded in order for it to be communicated, used, processed, and manipulated. If surrogate records and metadata statements are not converted to electronic form, they cannot be stored and retrieved in online retrieval tools or on the web. This chapter addresses several ways to encode metadata and surrogate records, including the MARC (MAchine-Readable Cataloging) format and various XML (Extensible Markup Language) schemas. It also provides a brief overview of a new approach to encoding being developed for the linked data landscape of the Semantic Web.

The MARC format is the encoding standard currently being used to create bibliographic records that are stored in most library catalogs. It is a mature standard that was designed in the mid-1960s, became an international standard in 1973, and has been in worldwide use for over four decades.[1] Although some call it a *mature* standard, others might refer to it as an *aging* standard. In the first decades of the twenty-first century, MARC has really begun to show its limitations, and many in the LIS professions support replacing it with something less proprietary and more compatible with the web technologies of today (and tomorrow).

Another standard, SGML (Standard Generalized Markup Language), also had its origins in the mid-1960s. It developed as a *meta-language* for encoding electronic texts. In other words, it is a markup language for creating markup languages; it provides a set of rules by which other more specific markup languages may be created. SGML became an international standard in 1986.[2] In addition to its standard function of marking up a wide variety of electronic documents, SGML and its by-products have been adapted for encoding surrogate records. Various applications of SGML have been created over the years, including HTML (Hypertext Markup Language) and XML. Specific DTDs (document type definitions) and XML schemas also have been designed to encode data for individual communities or for particular document types. All of these standards may be used to store, share, and retrieve metadata in retrieval tools or to manage and structure documents on the web or in other electronic environments. Although XML-based schemas are now fairly common, XML is not without its detractors, with some saying it is far too verbose and difficult to understand.[3]

In the minds of many information professionals, metadata content and encoding for that content are inextricably entwined. Metadata records can be created by first determining the descriptive content and then encoding the content manually afterward or by filling in the blanks in a metadata record template comprising the various codes. Each of the encoding standards discussed in this chapter acts as a kind of container. Although it is a simplification, it is sometimes helpful to think of encoding or syntax as a bucket or a box that does little more than hold the precious metadata content. It can be thought of as a way of carrying the metadata from one place to another. The metadata content held in the container may be guided by a separate descriptive standard, or in a few rare cases, by the encoding standard itself. For example, the metadata entered into a MARC record is usually controlled by outside descriptive cataloging rules (e.g., *RDA: Resource Description & Access*), by the conventions of a thesaurus (e.g., *Library of Congress Subject Headings*), and by the rules of a classification scheme (e.g., *Dewey Decimal Classification*); there are, however, some MARC 21 fields for which the rules for content appear *only* in the MARC 21 standard itself (e.g., the MARC 21 field 007 gives specific codes for certain materials, such as microforms or motion pictures). Although some metadata standards include both encoding and content specifications, in this book, encoding standards are discussed separately before covering the creation of metadata content. This is an attempt to show that creating content and encoding metadata are not the same processes. In other words, the same content can be coded with any one of several different encoding standards (e.g., the same metadata could be placed in a MARC record or in an XML-encoded document header). It is helpful to understand the shapes and sizes of containers before starting to fill them, so we will review some of the attributes of several encoding formats. Creation of metadata content is covered in Chapters 7–11.

ENCODING OF CHARACTERS

Before addressing the encoding of records and documents, it may be useful to say a few words about how computers are able to handle the very basic level of the characters that make up content that appears as a record. Each letter of an alphabet, each numeral, and each symbol in every language that exists has to be represented by a code, because, fundamentally, computers just deal with numbers. They store letters and other characters by assigning a code to each one. An International Organization for Standardization (ISO) standard, ISO/IEC 10646, *Universal Coded Character Set* (UCS) is the first standardized character set with the purpose of including all characters in all written languages of the world (including all mathematical and other symbols). Unicode is an American industry counterpart that is kept compatible with ISO/IEC 10646. Unicode is overseen by the Unicode Consortium and was named for its aim to embrace three characteristics: *universality,* covering all modern written languages; *uniqueness,* with no duplication of characters even if they appear in more than one language; and *uniformity,* with each character being the same length in bits. UCS/Unicode provides a unique code for every character in every modern language. Before UCS/Unicode, there were a multitude of different encoding systems for the display of characters, such as ASCII (American Standard Code for Information Interchange) and EBCDIC (Extended Binary Coded Decimal Interchange Code). In the latest version of Unicode (10.0), over 8,500 characters from four new scripts (including new signs, symbols, and 56 emojis) were added for a total of 136,690 characters available for use.[4]

ENCODING OF RECORDS

In order for metadata to be accessible, it must be conveyed to users through some type of information retrieval tool. Now that our retrieval tools are primarily online (at least in most of the industrialized world), metadata is usually created in electronic form. Although some print retrieval tools are still used, even those are usually created using electronic means. Metadata records and statements are encoded in electronic form by assigning tags, numbers, letters, or words (i.e., codes) to discrete pieces of information in the description. For example, a primary creator's personal name is given the tag 100 in MARC coding, or in an XML schema such as the Text Encoding Initiative (TEI), it is preceded with **<author>** and is followed by **</author>**. Such encoding is often referred to in metadata schemas as the *syntax* of the schema.

Surrogate records are encoded in order to be able to provide users access to the contents of those surrogate records. Encoding allows for the setting off of individual parts of the metadata (called *fields* or *elements*) for specific purposes. This enables the creation of software programs that allow the indexing and searching of only certain fields. For example, if a user wishes to search only for personal authors, the catalog will narrow the search to a stockpile of metadata that has been coded as either 100 or 700 in MARC-based metadata. As noted above, the same result is achieved in some XML schemas by placing a name between opening and closing **<author>**, **<creator>**, or **<contributor>** tags.

Encoding also is used to provide for display. In a similar manner to the searching of only certain fields, programs can be written so that each field will display in a certain position on the screen according to the wishes of those creating the display. When a TEI **<author>** tag or a MARC 100 or 700 field is used to identify the name of a person who is the creator of an information resource, that creator's name can be presented consistently in the system. For example, the tags used to identify the creator can then tell the system to display the name at the top of the screen or just after the title; to present the name in large boldface type or in italics; and to precede the name with a particular label or not.

Another use for encoding is to allow many languages and scripts to be displayed and searched in the same file. Languages that are in the Roman alphabet have always been able to be interfiled, although filing them alphabetically was sometimes a problem (see discussion of filing issues in Appendix C: Arrangement of Metadata Displays). Languages in other scripts had to be *Romanized* in order to interfile them in the paper world. Online, however, if the display of non-Roman scripts is provided for, record encoding allows for identification of the fields of the surrogate record regardless of the language or script.

Encoding also is used for data transmission. Institutions that collect information resources tend to have their own online systems into which they place encoded metadata for their resources. When these information resources are duplicates of those found in other institutions' online systems, cooperative arrangements are made to exchange the metadata. This *cooperative cataloging* means that each institution does not have to create a metadata description for every resource "from scratch." Shared, standardized encoding allows such exchanges of metadata to occur among institutions around the world.

There are a number of standards in use for encoding metadata. The examples discussed here are:

- MARC (MAchine-Readable Cataloging)
    - MARC 21
    - UNIMARC (Universal MARC)
- SGML (Standard Generalized Markup Language)
    - HTML (Hypertext Markup Language)
    - XML (Extensible Markup Language) and its DTDs and Schemas
        - TEI (Text Encoding Initiative) Schema
        - EAD (Encoded Archival Description) DTD and Schema
        - ONIX (Online Information Exchange) DTD and Schema
        - MARCXML Schema and MODS (Metadata Object Description Schema)

MARC (MACHINE-READABLE CATALOGING)

In the 1960s the Library of Congress (LC) created the MARC format as a way to share bibliographic data among libraries. The MARC communications format is used for encoding metadata and transmitting it from one system to another (see Figure 6.1). It is primarily used in libraries, although it has been used in some museums and archives over the years.

The MARC format comprises several components, including a leader, a directory, and two types of variable fields (control fields and data fields). The *leader* identifies the beginning of a new record and contains 24 alphanumeric characters that identify, among other things,

- record length,
- record status,
- character-coding scheme (e.g., UCS/Unicode),
- encoding level, and
- the kind of information resource that is being described.

04277cam a2200781 i 4500
001001300000003000600013005001700019008004100036010001700077040010000094019003600194020004700230
020004400277020004200321020003900363020002600402035007500428042000800503050002000511082001400531084003400545
084003200579084001800611049000900629100003300638245011100671250002200782264006200804300003000866336002600896
337002800922338002700950490004800977500015201025504006901177505094401246520038102190650002802571650002402599
650002702623630003502650650010402685650006102789630006102850650005302911650005402964650005003018630006103038
650004003129650004503169650004403214700003903258700005303297700009603350830004903446
ocn911180115OCoLC20160926172003.0150616s2015 cau b 001 0 eng a 2015012911 aDLCbengerdacDLC
dYDXdMBBdYDXCPdBDXdBTCTAdOCLCFdNTEdCDXdNDSdIBldCHVBKdGBVCPdOCLCO a881400756a881446762a935877570
a9781598848571 (hardback : acid-free paper) a1598848577 (hardback : acid-free paper) a9781598848564 (pbk : acid-free paper)
a1598848569 (pbk : acid-free paper) z9781440837456 (ebook) 1 a(OCoLC)911180115z(OCoLC)881400756z(OCoLC)881446762
z(OCoLC)935877570 apcc00aZ693b.W94 201500a025.3223 aLAN025030aLAN0200002bisacsh a06.709Katalogisierung2bcl aAN
740002rvk aDD0A1 aJoudrey, Daniel N.,eauthor.10aIntroduction to cataloging and classification /cDaniel N. Joudrey, Arlene G. Taylor,
and David P. Miller. aEleventh edition. 1aSanta Barbara, California :bLibraries Unlimited,c[2015] axxv, 1048 pages ;c27 cm.
atextbtxt2rdacontent aunmediatedbn2rdamedia avolumebnc2rdacarrier1 aLibrary and information science text series aSignificant
expansion of: Introduction to cataloging and classification. Tenth edition / Arlene G. Taylor ; with the assistance of David P. Miller.
aIncludes bibliographical references (pages 1001-1022) and index.0 aCataloging in context -- Development of catalogs and cataloging
codes -- Underlying principles and conceptual models -- Resource Description and Access (RDA) basics -- Manifestations and items --
Works and expressions -- Persons, families, places, and corporate bodies -- Relationships and the use of access points -- RDA metadata
in the MARC format -- Authority control -- Subject access -- Verbal subject access -- Library of Congress subject headings (LCSH) -- Sears
list of subject headings (Sears) -- Other verbal access systems -- Classification -- Decimal classification -- Library of Congress
classification (LCC) -- Creation of complete call numbers -- Other classification systems -- MAchine-Readable Cataloging (MARC)
encoding -- Alternative containers for metadata -- International standard bibliographic description (ISBD) -- Cataloging management
and support -- Appendix A: RDA outline -- Appendix B: ICC11 RDA book template. aThis new edition reintroduces the topic of library
cataloging from a modern perspective. This textbook delineates the new cataloging landscape; shares a principles-based perspective;
provides introductory text for beginners and intermediate students; emphasizes descriptive and subject cataloging, as well as format-
neutral cataloging, and covers new cataloging rules and RDA. 0aDescriptive cataloging. 0aSubject cataloging. 0aClassificationxBooks.
00aResource description & access. 7aLANGUAGE ARTS & DISCIPLINES / Library & Information Science / Cataloging & Classification.
2bisacsh 7aLANGUAGE ARTS & DISCIPLINES / Study & Teaching.2bisacsh07aResource description & access.2fast0(OCoLC)fst01791077
7aClassificationxBooks.2fast0(OCoLC)fst00863572 7aDescriptive cataloging.2fast0(OCoLC)fst00891123 7aSubject cataloging.
2fast0(OCoLC)fst0113645007aResource description and access.0(DE-588)7710221-62gnd 7aBibliothek.0(DE-588)4006439-62gnd
7aKatalogisierung.0(DE-588)4163418-42gnd 7aKlassifikation.0(DE-588)4030958-72gnd1 aTaylor, Arlene G.,d1941-eauthor.1 aMiller,
David P.q(David Peter),d1955-eauthor.1 iBased on (work):aTaylor, Arlene G.,d1941-tIntroduction to cataloging and classification.
0aLibrary and information science text series.

Figure 6.1. A MARC Bibliographic Record Displayed in the Communication Format. (Source: OCLC Connexion, WorldCat—Record Number 911180115)

The leader can be seen in Figure 6.1 (it is the top line in boldface type).

The *directory* contains a series of fixed-length segments (each containing 12 characters) that identify the field tag, length, and starting position of each data field in the record. The directory is visible in Figure 6.1, but it is the focus of Figure 6.2. In Figure 6.2, one segment is highlighted; it is conveying that the 700 field (an additional creator/contributor access point for a personal name) is 39 characters long and it begins at the 3,258th character position in the record. The MARC directory was a major contribution to information science. Instead of containing only fixed-length fields, which was common in databases in the 1960s, the directory structure allowed for variable-length fields, which are now prevalent.

The MARC format identifies two types of variable fields: variable control fields and variable data fields. *Control fields* carry alphanumeric data elements; they are used for processing machine-readable bibliographic records. In MARC 21 these field tags begin with two zeros (i.e., 00X fields). Control fields are used for such data as fixed-length descriptive data, LC control numbers, and codes for date and time of latest transaction. Some control fields are referred to as fixed fields, but one of these (the 008 field) is often called *the fixed field* and is displayed differently from other fields in some online systems.

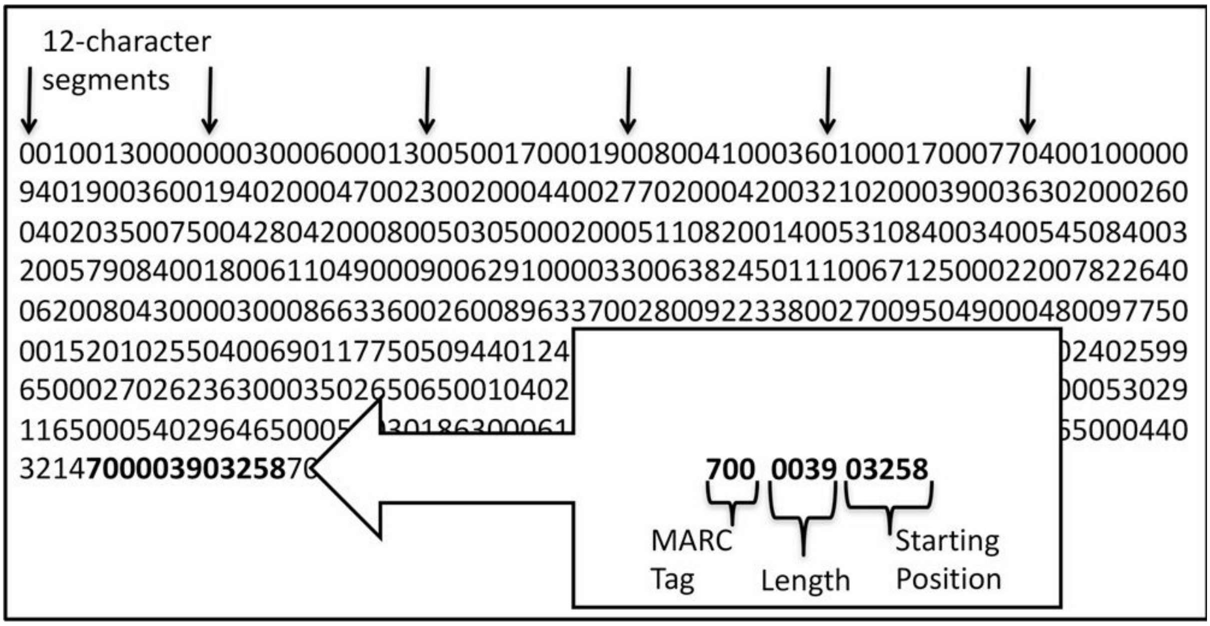


Figure 6.2. The Directory in a MARC Bibliographic Record. (Source: OCLC Connexion, WorldCat—Record Number 911180115)

Variable data fields carry alphanumeric data of variable length. These fields contain traditional cataloging data but also may carry information such as playing time, URLs, and linking entries. Variable data fields often have quite a few subfields in order to achieve greater granularity in encoding. Some commonly used variable fields include

- 020 for International Standard Book Numbers (ISBNs)
- 082 for Dewey Decimal Classification (DDC) numbers
- 100 for primary creator (personal name)
- 245 for title statements and statements of responsibility
- 264 for dissemination information (publication, manufacture, and distribution statements)
- 300 for descriptions of the physical formats of resources
- 490 for series statements
- 500 for general notes
- 650 for subject headings
- 655 for genre/form headings

Some variable fields are repeatable in a single record (e.g., 650 fields containing subject headings) and others are not repeatable (e.g., a 100 field containing the primary creator)—it all depends on the purpose of the field.

Variable data fields are made up of *tags*, *indicators*, and *subfield codes*. *Tags* are the three-digit numbers (from 001 to 999) that designate the kind of content that will be entered into the field. Although there is a potential to have as many as 999 MARC tags, many fewer (nearer to 275) are actually defined, and even fewer are frequently used. In 2010 an OCLC report showed that only 11 MARC tags were found in 20 percent or more of their 145 million bibliographic records, and that 22 tags occurred in 10 percent or more of those records.[5] Of the available MARC tags, clearly most are used only rarely. LIS researchers William Moen and Penelope Benardino presented similar findings in 2003.[6] The 22 most frequently used MARC tags are listed in Table 6.1.

When records are received in the MARC communications format, each system displays them according to its own programming. OCLC (Online Computer Library Center in Dublin, Ohio) puts the fixed field at the top and uses mnemonic abbreviations to interpret the various codes that make up the 008 field. LC's catalog displays the fixed field as a long string of alphanumeric characters in the 008 position without abbreviations or separations. Local systems have a staff display that is quite different from the network systems' displays. The public interface to the catalog generally does not display the fixed field at all (except in implementations that allow for a MARC view). System displays also vary in such things as spacing, placement of certain subfields, and so on (see Figures 6.3 and 6.4).

A convention is followed in which all MARC fields beginning with the same first digit are identified as a group. The groups are

1XX	Main entry fields
2XX	Title and edition fields
3XX	Physical description fields
4XX	Series fields
5XX	Notes fields
6XX	Subject fields
7XX	Added entry and linking fields
8XX	Series added entry and holdings data fields
9XX	Local data fields

Table 6.1. The Most Frequently Used MARC Tags in OCLC Bibliographic Records.

Tag	Description	Example	
001	**Control Number**	**001**	**18660797**
007	**Physical Description Fixed Field**	**007**	**sd fungnnmmneu**
008	**Fixed Field**	**008**	**150616s2015 cau b 001 0 eng**
010	**Library of Congress Control Number**	**010**	**$a 2015012911**
016	**National Bibliographic Agency Control Number**	**016 7_**	**$a FRBNF443440320000006 ‡2 FrPBN**
020	**International Standard Book Number**	**020**	**$a 9781598848564**
040	**Cataloging Source**	**040**	**$a DLC $b eng $e rda $c DLC $d YDX**
043	**Geographic Area Code**	**043**	**$a n-us-ma**
050	**Library of Congress Call Number**	**050 00**	**$a Z693 $b.W94 2015**
082	**Dewey Decimal Classification Number**	**082 00**	**$a 025.3 $2 23**
100	**Main Entry – Personal Name**	**100 1_**	**$a Joudrey, Daniel N., $e author.**
245	**Title Statement**	**245 10**	**$a Introduction to cataloging and classification / $c Daniel N. Joudrey, Arlene G. Taylor, and David P. Miller.**
250	**Edition Statement**	**250**	**$a Eleventh edition.**
260	**Imprint Statement** ***Note: In more recent RDA-compliant records, the 260 is replaced with a 264 field.***	**264 _1**	**$a Santa Barbara, California :** **$b Libraries Unlimited, $c [2015]**
300	**Physical Description**	**300**	**$a xxv, 1048 pages : $b illustrations; $c 27 cm**
440	**Series Statement/Added entry [Obsolete]** ***Note: Replaced by a 490/8XX combination.***	**830 _0**	**$a Library and information science text series.**
490	**Series Statement**	**490 1_**	**$a Library and information science text series**
500	**General Note**	**500**	**$a Significant expansion of: Intro- duction to cataloging and classifi- cation. Tenth edition / Arlene G. Taylor ; with the assistance of David P. Miller.**
504	**Bibliography, etc. note**	**504**	**$a Includes bibliographical references (pages 1001–1022) and index.**
650	**Subject Added Entry – Topical Term**	**650 _0**	**$a Subject cataloging.**
700	**Added Entry – Personal Name**	**700 1_**	**$a Taylor, Arlene G., $d 1941- $e author.**
710	**Added Entry – Corporate Name**	**710 2_**	**$a Association for Library Collections & Technical Services.** **$b Subject Analysis Committee,** **$e sponsoring body.**

OCLC 911180115 No holdings in DD0 - 190 other holdings
No DD0 holdings in GLIMIR cluster; 195 other holdings in GLIMIR cluster

Books		**Rec Stat** c		**Entered** 20150616			**Replaced** 20170609083829.5					
	Type	a	**ELvl**		**Srce**		**Audn**		**Ctrl**		**Lang**	eng
	BLvl	m	**Form**		**Conf**	0	**Biog**		**MRec**		**Ctry**	cau
			Cont	b	**GPub**		**LitF**	0	**Indx**	1		
	Desc	i	**Ills**	a	**Fest**	0	**DtSt**	s	**Dates**	2015,		

010 2015012911
040 DLC **‡b** eng **‡e** rda **‡c** DLC **‡d** YDX **‡d** MBB **‡d** YDXCP ...
020 9781598848571 (hardback : acid-free paper)
020 9781598848564 (pbk : acid-free paper)
020 **‡z** 9781440837456 (ebook)
050 00 Z693 **‡b** .W94 2015
082 00 025.3 **‡2** 23
100 1_ Joudrey, Daniel N., **‡e** author.
245 10 Introduction to cataloging and classification / **‡c** Daniel N. Joudrey, Arlene G. Taylor, and David P. Miller.
250 Eleventh edition.
264 _1 Santa Barbara, California : **‡b** Libraries Unlimited, **‡c** [2015]
300 xxv, 1048 pages : **‡b** illustrations ; **‡c** 27 cm.
336 text **‡b** txt **‡2** rdacontent
337 unmediated **‡b** n **‡2** rdamedia
338 volume **‡b** nc **‡2** rdacarrier
490 1_ Library and information science text series
500 Significant expansion of: Introduction to cataloging and classification. Tenth edition / Arlene G. Taylor ; with the assistance of David P. Miller.
504 Includes bibliographical references (pages 1001-1022) and index.
505 0_ Cataloging in context -- Development of catalogs and cataloging codes -- Underlying principles and conceptual models -- Resource Description and Access (RDA) basics -- Manifestations and items ... [*Table of contents continues*] ... Appendix B: ICC11 RDA book template.
520 This new edition reintroduces the topic of library cataloging from a modern perspective. This textbook delineates the new cataloging landscape; shares a principles-based perspective; provides introductory text for beginners and intermediate students; emphasizes descriptive and subject cataloging, as well as format-neutral cataloging, and covers new cataloging rules and RDA.
650 _0 Descriptive cataloging.
650 _0 Subject cataloging.
650 _0 Classification **‡x** Books.
630 00 Resource description & access.
700 1_ Taylor, Arlene G., **‡d** 1941- **‡e** author.
700 1_ Miller, David P. **‡q** (David Peter), **‡d** 1955- **‡e** author.
700 1_ **‡i** Based on (work): **‡a** Taylor, Arlene G., **‡d** 1941- **‡t** Introduction to cataloging and classification.
830 _0 Library and information science text series.

Figure 6.3. A Partial MARC Record as Displayed by OCLC. (Source: OCLC Connexion, WorldCat—Record Number 911180115)

```
000  02384cam a2200457 i 4500
001  18660797
005  20160818094544.0
008  150616s2015    cau     b    001 0 eng
010  __   |a 2015012911
020  __   |a 9781598848571 (hardback : acid-free paper)
020  __   |a 9781598848564 (pbk : acid-free paper)
020  __   |z 9781440837456 (ebook)
040  __   |a DLC |b eng |c DLC |e rda |d DLC
050  00   |a Z693 |b .W94 2015
082  00   |a 025.3 |2 23
100  1_   |a Joudrey, Daniel N., |e author.
245  10   |a Introduction to cataloging and classification / |c Daniel N. Joudrey, Arlene G.
               Taylor, and David P. Miller.
250  __   |a Eleventh edition.
264  _1   |a Santa Barbara, California : |b Libraries Unlimited, |c [2015]
300  __   |a xxv, 1048 pages : |b illustrations ; |c 27 cm.
336  __   |a text |b txt |2 rdacontent
337  __   |a unmediated |b n |2 rdamedia
338  __   |a volume |b nc |2 rdacarrier
490  0_   |a Library and information science text series
500  __   |a Significant expansion of: Introduction to cataloging and classification. Tenth
               edition / Arlene G. Taylor ; with the assistance of David P. Miller.
504  __   |a Includes bibliographical references (pages 1001-1022) and index.
650  _0   |a Descriptive cataloging.
650  _0   |a Subject cataloging.
650  _0   |a Classification |x Books.
630  00   |a Resource description & access.
700  1_   |a Taylor, Arlene G., |d 1941- |e author.
700  1_   |a Miller, David P. |q (David Peter), |d 1955- |e author.
700  1_   |a Taylor, Arlene G., |d 1941- |t Introduction to cataloging and classification. |i
               Based on (work):
```

Figure 6.4. A Partial MARC Record as Displayed in the Library of Congress Catalog. (Source: Library of Congress Online Catalog—Record Number 18660797)

These groupings mean, for example, that all 1XX fields are related to the primary creator of a resource and all 6XX fields are somehow related to subject access (e.g., 600 is for a personal name used as subject; 610 is for a corporate name used as subject; 650 is for a topical subject).

Indicators consist of two character positions following a tag. These positions contain coded information interpreting or supplementing the data in the field. The number of indicators used and the meaning of each indicator depend on the individual tag. For some fields, no indicators are used, so the indicator positions are empty. In other fields, one or two of the indicators may be defined and used. Each indicator position is independent of the other. An example in MARC 21 is the indicators for the title field. The first indicator in the 245 tells us whether a title added entry is needed.[7] The second indicator tells us how many characters should be skipped over (counting a blank space as one character) to get to the first word in the title that is *not* an article.

245 14 $a The organization of knowledge in libraries : $b and the subject-approach to books / $c by Henry Evelyn Bliss.

In this example, the second indicator is *4* because the title begins with the article *the* and we want the title to be arranged with *O* titles as *Organization*, not with *T* titles as *The Organization*.

A *subfield code* consists of a delimiter and a letter or number. The purpose of subfield codes is to introduce a greater level of granularity into MARC encoding. They identify elements in a field that might require separate treatment; they break down the larger field into smaller, more manageable chunks. Some subfields may be repeatable to accommodate multiple values for a particular element (e.g., two publication locations); repeatability depends on the specific MARC field and the purpose of the individual subfield.

Delimiters are unique characters that indicate the beginning of a particular subfield. The delimiter may be represented differently depending on the system (e.g., $ or | or ‡). The lowercase letter or number that follows the delimiter is a *data element identifier*. It specifies the content found in the subfield. For example, in the 245 example above, subfield *a* ($a) identifies the *title proper* (the resource's main title), $b identifies *other title information* (e.g., subtitles, parallel titles), and $*c* identifies the *statement of responsibility* (e.g., the creators and others responsible for this resource). The subfield letters and numbers, however, have different meanings from field to field. For example, $b in the 250 field contains the remainder of an edition statement and $b in the 110 field holds a subordinate unit that is part of the corporate body named in $a. (It should be noted that in some systems, most notably OCLC, the $a code of the first element in a field is not shown by the system.) These and other key components of a MARC record are identified in Figure 6.5.

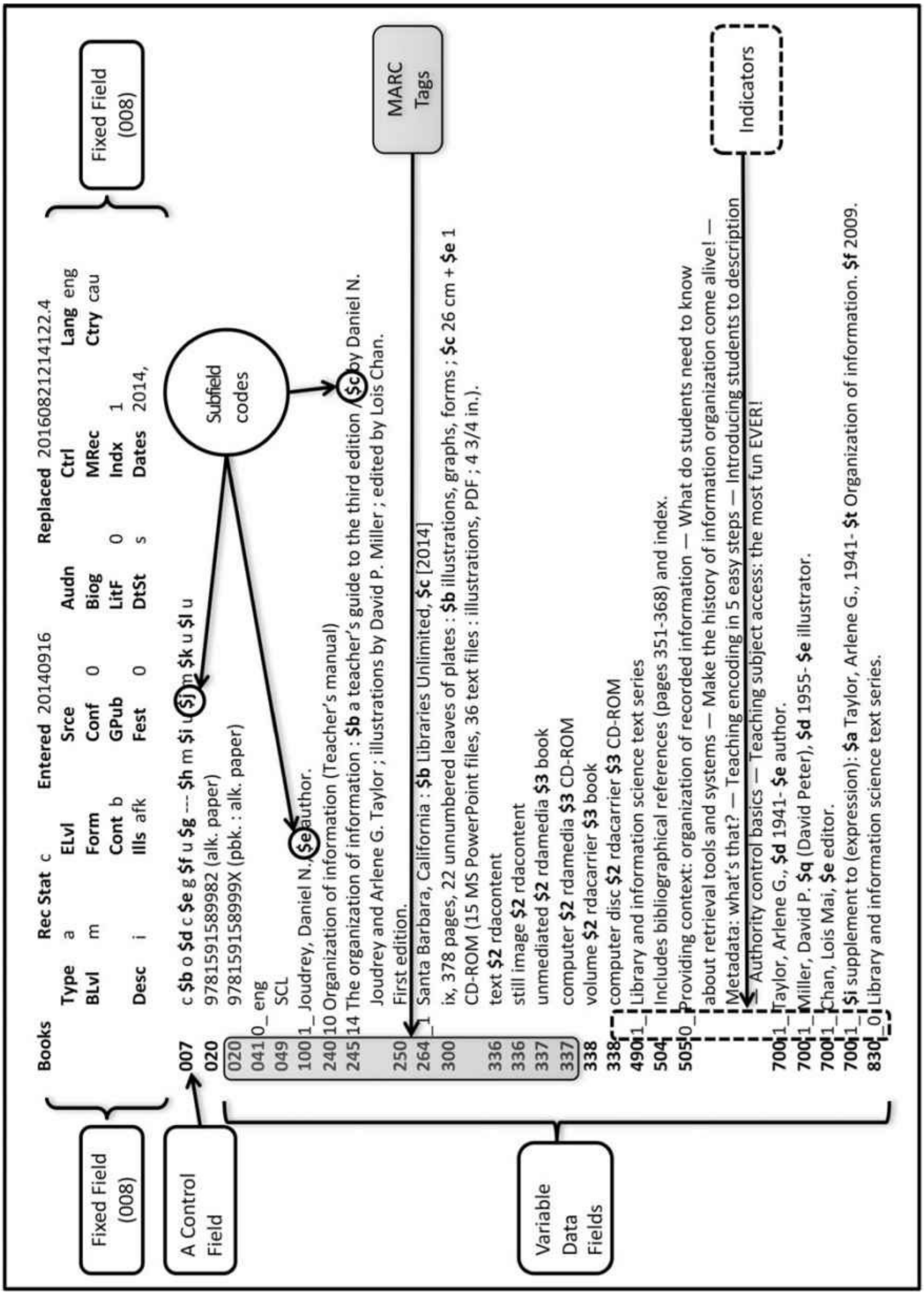


Figure 6.5. A MARC Record with Components Highlighted. (Display Based on the OCLC Format)

MARC 21

MARC 21 (named *21* for the twenty-first century) came into being in 2000 as a result of agreement between the United States and Canada to merge their national MARC formats (USMARC and CAN/MARC), compromising on the differences between them. In 2004 the British Library adopted MARC 21 as its cataloging format, and that same year, the German National Library announced it also would be moving to MARC 21. The first MARC format was developed at LC in 1968 under the leadership of Henriette Avram. It was identified simply as MARC until other versions were developed in the 1970s, and then the first version was referred to as LC-MARC. Finally, the name USMARC came to distinguish it from more than 20 other national versions (e.g., UKMARC, RUSMARC, IBERMARC).

MARC 21 is based on ANSI/NISO standard Z39.2-1994, *Information Interchange Format* (originally the *American National Standard for Bibliographic Information Interchange on Magnetic Tape* published in 1971, revised in 1979, and again in 1984).[8] The international version is ISO 2709:2008, *Information and Documentation—Format for Information Exchange.*[9] MARC 21 records may be encoded with either an 8-bit encoding called MARC-8 or with the Unicode UTF–8 encoding rules.[10]

MARC 21 actually has five formats for different types of data. These include the following:

- **Bibliographic format**—for encoding bibliographic data in records that are surrogates for information resources (see Figures 6.3–6.4)
- **Authority format**—for encoding authority data collected in authority records created to help control the content of those surrogate record fields that are subject to authority control (see Figure 8.3 on page 380)
- **Community information format**—for encoding data in records that contain information about events, programs, services, and the like, so that these records can be integrated into catalogs
- **Holdings format**—for encoding data elements in holdings records that show the holdings and location data for information resources described in surrogate records
- **Classification data format**—for encoding data elements related to classification numbers, the captions associated with them, their hierarchies, and the subject headings with which they correlate

For a more complete description of MARC 21, please see Daniel N. Joudrey, Arlene G. Taylor, and David P. Miller's 11th edition of *Introduction to Cataloging and Classification*.[11]

UNIMARC

UNIMARC (Universal MARC)[12] was developed in 1977 as a vehicle for interchange of MARC records between national bibliographic agencies. The most recent version includes updates made in 2012. It conforms to ISO 2709, as does MARC 21. UNIMARC calls for use of ISO character set standards but also allows parties to agree on the character set to be used when exchanging records.

The proliferation of national MARC formats necessitated development of UNIMARC. At first it was thought that UNIMARC would act only as a conversion format. In this capacity, it requires that each national agency create a translator to change records from UNIMARC to the particular national format and vice versa. When a translator is in place, records can be converted to UNIMARC to be sent to other countries, and records received from other countries can be translated from UNIMARC to the national format. In addition to this use, a few national agencies that did not already have a MARC format have adopted UNIMARC as the standard in their countries.

Differences between UNIMARC and MARC 21 are immediately apparent upon looking at the list of code blocks beginning with 0, 1, 2, etc. For example, the 1XX fields are *coded information* rather than *main entry* fields, and the 2XX block is a *descriptive information* block in which a title field is designated 200, rather than 245 as in MARC 21. Also, the second indicator of the 200 field is blank and therefore does not tell how many characters to skip for an initial article. UNIMARC also allows for embedded fields (e.g., an authority record number may appear at the beginning of an access point field). It de-emphasizes main entry in the 7XX block, where 700 is for Personal Name–Primary Intellectual Responsibility, but if the concept of main entry does not exist or is not distinguished in a source format, then 701 is used for all personal names.

The Future of MARC

In recent years, some information professionals—particularly those working with nonprint resources—have expressed dissatisfaction with the limitations of the MARC format. They have suggested that librarians should be developing another encoding format for its bibliographic metadata in order to remain current and relevant in today's web-based information environment. Others (primarily catalogers) have defended the MARC standard, stating that it has more than met its purpose over the years and continues to be relevant. As with most standards, there are both strengths and weaknesses associated with MARC. Strengths include the fact that MARC is a mature standard. It has been in existence for over five decades, and it is well understood and well tested. It has been widely adopted by libraries around the world. Millions of records have been created in the last 50 years and are available to libraries through bibliographic networks. MARC has long been established as the format used in the typical integrated library system (ILS), and as of yet, few ILSs process information in other encoding formats (see the discussion of ILSs in Chapter 4).

Weaknesses include the fact that MARC is virtually unknown outside of libraries. This limits the ability of non-MARC information institutions to use the bibliographic data created by libraries. Some believe this marginalizes librarians and library metadata in the wider web-based information environment. Others find MARC's size limitations and its inability to convey hierarchical or complex relationships among entities to be problematic. Some dislike MARC's inability to embed related objects in the record (e.g., book covers), and see this as symptomatic of MARC's antiquated data structure.

Some, for many years, suggested that an XML schema should replace MARC. Others have declared that instead of arguing over MARC versus XML, we should be discussing MARC *and* XML in order to take advantage of the interoperability of XML without giving up the established content designation structure of MARC (see discussion of MARCXML below). In more recent years, some have suggested that libraries need to encode bibliographic metadata in a linked data format for use outside the silo of the ILS. LC is actively pursuing this direction in its Bibliographic Framework Initiation (see discussion of BIBFRAME below). At the time of this writing, the future of MARC seems assured for at least the next decade; but as technology continues to advance, MARC's age and limitations become more apparent.

SGML (STANDARD GENERALIZED MARKUP LANGUAGE)

The Standard Generalized Markup Language (SGML) is a *meta-language*—that is, a language for describing markup languages. Markup languages are used to describe the structure and contents of documents. Its origins can be traced back to work done at IBM. In 1969 Charles Goldfarb, Ed Mosher, and Ray Lorie developed what was at first named the *Text Description Language* but was eventually called *Generalized Markup Language* (GML).[13] In the 1980s it continued to evolve until it was published as an international standard: ISO 8879:1986, *Information Processing—Text and Office Systems—Standard Generalized Markup Language* (SGML).[14]

Unlike the MARC family of encoding languages, however, SGML does not contain a predefined set of tags that can be used to mark up documents, nor is it a standard template for producing particular types of documents. Instead, it is a set of guidelines for designing hardware-independent markup languages. Markup languages describe the structures of documents so that documents may be interchanged across computer platforms. Both text and markup are encoded in the same character set (e.g., Unicode or ASCII text), but SGML allows documents to be represented in such a way that the document structure may be identified independently from the content. *Document structure* means coding that says: this piece of text is the title, this content is the first chapter, this string of characters is a section heading, these lines comprise a paragraph, this is a quoted statement, and so on. Or, for surrogate records, the coding says this is a title element, this is a creator element, this is a publisher element, and so forth. It is said that SGML is flexible enough to define an infinite number of markup languages.

SGML defines data in terms of entities, elements, and attributes. An *entity* is a thing to be encoded (e.g., a document, a part of a document, a surrogate record). An *element* is a particular component in the entity, such as a title, a chapter title, a section heading, a paragraph, a publisher name, a classification number, and so forth. An *attribute* provides particular information about an element (e.g., the type of place, the language used in the element). The relationships between entities, elements, and attributes are described using SGML.

SGML prescribes markup that consists of delimiters and tags. Delimiters are defined symbols (e.g., <, >, </, "), and are used to construct tags (e.g., **<author>** is a tag). Tags usually appear before and after an element. For example, if an element (such as an author's name) is to be coded, then that name is placed between two tags identifying the type of element:

Basic Structure: <**tag**>element</**tag**>
Example: <**author**>Edward Gaynor</**author**>

Attribute values are delimited by using quotation marks (". . ." or '. . .').

<**quote lang**='spa'>¿Que pasa?</**quote**>

Tags may be nested hierarchically. An example of a nested set of tags from TEI (described below) is found in Figure 6.6.

```
<teiHeader>
    <fileDesc>
        <titleStmt>
            <title>Introduction to Cataloging and Classification</title>
            <author>Daniel N. Joudrey</author>
            <author>Arlene G. Taylor</author>
            <author>David P. Miller</author>
            <respStmt>
                <resp>Encoded by</resp>
                <name>Amy Rudersdorf</name>
            </respStmt>
        </titleStmt>
        ...
    </fileDesc>
</teiHeader>
```

Figure 6.6. An Example of Nesting Tags in a TEI Header. (TEI coding shown in bold)

Because SGML does not contain a prescribed set of tags, it requires some form of structure to be provided through a particular application or through a DTD at the beginning of a document (or to declare a particular external, pre-established DTD as the one being followed). DTDs are discussed further below. Before discussing DTDs, however, two specific applications or derivations of SGML are discussed: HTML and XML.

HTML (Hypertext Markup Language)

HTML, originally a derivation of SGML, was developed for the creation of web pages.[15] It is a basic markup language that allows almost anyone to be a web author. It defines the content, the layout, and the formatting of web documents. It provides for creation of a simple structure, enables display of images, and provides for establishment of links between documents. Users of an HTML-encoded document can navigate through the text itself if internal links have been made, or can move from one website to another with external links. In HTML there are specific provisions for elements that can be used to describe properties of a document (e.g., title, date, keywords), making it possible to encode rudimentary metadata. Extensible Hypertext Markup Language (XHTML)[16] is a reformulation of HTML4 as an XML application. A major benefit of migrating to XHTML is in being conformant to XML, while still having the content be able to be displayed correctly by older browsers. In 2014 HTML5 was released by the World Wide Web Consortium (W3C) to replace both HTML4 and XHTML and to decrease the reliance on plug-ins, such as Flash and Java, when providing access to multimedia content.[17] HTML5, however, is no longer compatible with SGML.

Until recently, it was rare to find metadata encoded in HTML. In the early years of the web, the creation and insertion of *meta tags* in the headers of HTML documents was encouraged by the W3C. Although specific metadata elements were not defined in the HTML standard, page creators could use any metadata element set as long as a uniform resource identifier (URI) could be provided for that set in the generic meta tags. The Dublin Core Metadata Initiative also encouraged web designers to insert Dublin Core metadata into their document headers as part of the creation process. Much to the disappointment of many information professionals, however, the make-your-own-metadata approach never really caught on with most creators of web content; or when it did, it often involved mischief or malfeasance (e.g., attempts to fool search engines into directing users to an irrelevant page through misrepresentations of the content). By 2002 no existing search engine gave any credence to the keyword meta tags found in web documents.[18]

More recently, however, major search engines are beginning to trust metadata annotations that are being embedded within the HTML coding of web documents. Annotations may come in a variety of forms. For example, some may use HTML microdata[19] to annotate the content of a resource, but others may prefer to use microformats2,[20] RDFa (Resource Description Framework in Attributes),[21] or another form. The simplest format, microdata, is an HTML5 specification that can be used to nest metadata, in the form of name-value pairs, within the content of web resources. Annotations must be made using a controlled vocabulary to provide semantic structure. Some annotation systems, such as microdata, depend on external vocabularies to provide the semantics (e.g., it can use the Schema.org vocabulary), but others, such as microformats2, have developed their own semantic structures for specific information types commonly found on the web. For example, to describe a recipe using microdata, you can use the 11 properties provided by Schema.org for that type of information.

- cookTime
- cookingMethod
- nutrition
- prepTime
- recipeCategory
- recipeCuisine
- recipeIngredient
- recipeInstructions
- recipeYield
- suitableForDiet
- totalTime

Schema.org also acknowledges that many of the properties associated with describing *creative works* and other categories of resources may also be needed (e.g., *name* for the title of the recipe). No matter which annotation format is used, the goal is still the same: to embed additional structure, context, and metadata into the HTML coding of web documents to enhance the efficiency of search engines. An example of some very simple HTML is presented in Figure 6.7.

```
<html>
<head>
<title>LIS 417 Syllabus</title>
</head>
<body itemscope itemtype="http://schema.org/WebPage">
...
<h1 align="center" itemprop="name" >LIS 417:  Subject Cataloging and Classification</h1>
<h2 align="center" itemprop="description">Syllabus</h2>
<div itemscope itemtype="http://schema.org/Person">
<ul>
<li><b>Instructor: </b><span itemprop="name">Daniel N. Joudrey</span>
<li><b>Title: </b><span itemprop="jobTitle">Associate Professor</span>
<li><b>Email: </b><span itemprop="email">joudrey@simmons.edu</span>
<li><b>Affiliation: </b><span itemprop="affiliation">School of Library and Information Science, Simmons
College</span></div><br><br>
<img src="me.jpg" itemprop="image" alt="Photo of DNJ"/>
</ul>
<br>
<br>
<div itemprop="address" itemscope itemtype="http://schema.org/PostalAddress">
        <span itemprop="streetAddress">300 The Fenway P-205B</span><br>
        <span itemprop="addressLocality">Boston</span>,
        <span itemprop="addressRegion">MA</span>
        <span itemprop="postalCode">02115</span>
        <span itemprop="telephone">(617) 555-5555</span>
</div>
<br>
<br>
<div itemscope itemtype="http://schema.org/CreativeWork">
<h2>Course Description: </h2>
<p itemprop="description"> This course addresses the theories, principles, and practices of subject
cataloging and classification. It covers the application of national standards to the creation of
bibliographic records and to the construction of catalogs in libraries and other information environments.
<br>
<br>
...
                        <!-- SYLLABUS goes on and on and on -- >

</body>
</html>
```

Figure 6.7. Simple HTML with Microdata Added. (HTML coding shown in bold; microdata elements shown in italics)

Over the years, as the web has grown and become more sophisticated in the kinds of activities and information that is stored there, HTML has been criticized as being too simplistic because it focused (at least initially) on display aspects and because it is not particularly good at representing complex document structures. To remedy this, some believed that full support of SGML on the web was the answer. SGML, however, is fairly complex and has features that make the programming involved complicated and lengthy. XML, a subset of SGML, was developed to address some of these concerns.

XML (Extensible Markup Language)

XML is a markup language used for web and other electronic documents, which can be read by both humans and machines. It is used to structure data and is touted by its developers as being just as easy to use as

HTML but at the same time being as powerful as SGML.[22] Unlike basic HTML, XML focuses on structural markup and leaves issues of display to style sheets instead. It also allows developers to create their own sets of tags, as needed—thus the name *Extensible* Markup Language. In addition, because XML was developed some years after SGML, it incorporates techniques needed for multimedia files, such as the ability to identify the format used to encode an illustration. It is essentially a version of SGML that can be used on the web. Its components are the same: entities, elements, and attributes. XML, however, allows for, but does not require, a document type definition (DTD).

A DTD is an XML or SGML application that defines, with its own notation, the structure of a particular *type* of document. It gives advance notice of what names and structures can be used in a particular document type, so that all documents that belong to a particular type will be alike. It could be thought of as a template for a particular type of document. Typically, the following are defined in a DTD:

- the elements that might be part of that particular document type
- element names and whether they are repeatable
- the contents of elements (in a general way, not specifically)
- what can be omitted
- tag attributes and default values
- names of permissible entities

For example, a DTD for a memo might have the following lines:

```
<!DOCTYPE memo [
<!ELEMENT memo (to, from, date, subject?, para+) >
<!ELEMENT to (#PCDATA) >
<!ELEMENT from (#PCDATA) >
<!ELEMENT date (#PCDATA) >
<!ELEMENT subject (#PCDATA) >
<!ELEMENT para (#PCDATA) >
] >
```

This DTD is stating that the document type is a *memo*, and that a memo includes a *to* element, a *from* element, and a *date* element; it includes one or more paragraphs, as indicated by the plus sign; and it may or may not include a *subject* element, as per the question mark. It also states that the *to*, *from*, and the other elements are made up of alphanumeric character strings (#PCDATA).

In the XML environment, there is an alternative to the DTD known as an *XML schema.* XML schemas have evolved as richer forms of DTDs, which define the content and semantics of documents, in addition to their structure. An XML schema differs from a DTD in being expressed in the XML syntax itself and in following XML rules. Therefore, a developer does not have to learn another notation, and the software does not have to have a different parser. An XML schema can also define shared vocabularies with links to the namespaces in which those vocabularies reside, and which cannot be part of a DTD. In large part, schemas have replaced DTDs because of these and other enhancements. In short, an XML schema defines:

- elements and attributes that can appear in a document
- parent-child relationships among elements
- the order and number of child elements
- whether an element is empty or can include text
- data types for elements and attributes
- default and fixed values for elements and attributes[23]

An XML schema for a memo would be written in XML and look like XML tagging:

```
<xs:element name="memo">
    <xs:complexType>
        <xs:sequence>
            <xs:element name="from" type="xs:string"/>
            <xs:element name="to" type="xs:string"/>
            <xs:element name="date" type="xs:string"/>
            <xs:element name="subject" minOccurs="0"
              maxOccurs="1" type="xs:string"/>
            <xs:element name="para" minOccurs="1"
              maxOccurs="unbounded" type="xs:string"/>
        </xs:sequence>
    </xs:complexType>
</xs:element>
```

This XML schema is stating that this document is a *memo;* it is made up of multiple elements in a particular sequence, which includes elements such as *from, to*, and *date*, which are made up of alphanumeric character strings.

An XML schema (or DTD) may be created for only one document, in which case it may be contained at the beginning of the text, but creating a schema is time consuming. It makes more sense to create schemas that can be used for many documents. These exist separately from the texts that refer to them. Many XML schemas have been created and are in general use. A few of these are discussed below as examples:

- **TEI Schema—**for encoding literary texts
- **EAD Schema—**for encoding archival finding aids
- **ONIX DTD** and **Schema—**for encoding publishers' records
- **MARCXML Schema** and **MODS—**for encoding MARC 21 records in XML

TEI (Text Encoding Initiative) Schema

When it was created in the late 1980s, TEI was designed as an SGML DTD, but over the years it has evolved first into an XML DTD, and then in the latest version, TEI 5, into an XML schema.[24] TEI was created to overcome the difficulty of having multiple encoding schemes being used to encode old, literary, and/or scholarly texts. Once encoded, the documents could not be exchanged easily. TEI makes features of a text explicit in a format that allows the processing of that text by different programs running on different machines. The text can be represented exactly as it appears in its original printed form (in the case of encoding text from books). Texts can be exchanged for research purposes (e.g., textual analysis). TEI can also be used for newly created documents, especially in cases where authors have a particular vision of how the text should look. TEI was originally created for texts in the humanities but is no longer limited to those texts. The guidelines provide a framework that can be used to describe many kinds of texts.

> The impact of the TEI on digital scholarship has been enormous. Today, the TEI is internationally recognized as a critically important tool, both for the long-term preservation of electronic data, and as a means of supporting effective usage of such data in many subject areas. It is the encoding scheme of choice for the production of critical and scholarly editions of literary texts, for scholarly reference works and large linguistic corpora, and for the management and production of detailed metadata associated with electronic text and cultural heritage collections of many types.[25]

TEI is intended to mark up a wide variety of materials for numerous user communities. By necessity, the set of tags included in the schema must be large, and may be considered unwieldy to some. To facilitate the ease of use of TEI, the consortium has designed particular *customizations* (i.e., subsets of the schema). These include tag sets specially designed for speech representations, dramas, manuscripts, journal articles, and the like. The most widely used of these customizations is TEI Lite.[26] It is a subset of TEI in much the same way that XML is a subset of SGML. TEI Lite contains a core set of tags from TEI to allow for basic encoding.

One particular component of TEI is of special interest to those who organize information: the *TEI Header*. It provides descriptive metadata about the electronic version of the resource encoded in the TEI document and the original source document, as well as administrative metadata about the encoding process, revisions, and so on. The TEI Header is discussed in Chapter 7. A partial example of a TEI Header containing metadata is provided in Figure 6.8. More complete examples may be found at the TEI website.[27]

```
<TEI version="5.0" xmlns="http://www.tei-c.org/ns/1.0">
<teiHeader>
    <fileDesc>
      <titleStmt>
            <title>Introduction to Cataloging and Classification</title>
            <author>Daniel N. Joudrey</author>
            <author>Arlene G. Taylor</author>
            <author>David P. Miller</author>
            <respStmt>
                  <resp>Digitized and encoded in TEI 5.0 by</resp>
                  <name>Amy Rudersdorf</name>
            </respStmt>
      </titleStmt>
      <editionStmt>
            <edition n="1">First digital edition</edition>
            <respStmt>
                  <resp>Proofed and edited by</resp>
                  <name>Emily Baldoni</name>
                  <name>Monica Shin</name>
            </respStmt>
      </editionStmt>
      <extent>
            <measure unit="Mb" quantity="132.5">132.5 megabytes</measure>
            <measure unit="pages" quantity="1048">1048 pages</measure>
      </extent>
      <publicationStmt>
            <pubPlace>Santa Barbara, CA</pubPlace>
            <publisher>ABC-CLIO eBooks</publisher>
            <date>2015</date>
            <idno type="ISBN">978-1-44083-745-6</idno>
            <availability status="restricted">
                  <p>Available under license from the publishers</p>
            </availability>
      </publicationStmt >
      <seriesStmt>
            <title>Library and Information Science Text series</title>
      </seriesStmt>
      <notesStmt>
            <note>Significant expansion of: Introduction to cataloging and classification. Tenth
            edition / Arlene G. Taylor ; with the assistance of David P. Miller</note>
      </notesStmt>
      <sourceDesc>
            <p>This first digital edition is based on the 11th edition of <title>Introduction to
            Cataloging and Classification</title> published in <date>2015</date> by Libraries
            Unlimited, an imprint of ABC-Clio</p>
      </sourceDesc>
    </fileDesc>
    ...
</teiHeader>
                        <!—A full TEI-encoded document follows the TEI Header -->
```

Figure 6.8. The File Description from a TEI Header. TEI encoding shown in bold)

EAD (Encoded Archival Description) Schema

EAD[28] is an XML DTD or schema for encoding finding aids used in archives and libraries. As with other DTDs and schemas, EAD does not specify intellectual content but defines the encoding designations. It eases the ability to exchange finding aids among institutions and allows users to find out about collections in distant places. Version 2002 of the EAD DTD functions both as an SGML and an XML DTD, as well as an XML schema. It conforms to all SGML/XML specifications. The minimum set of EAD tags required for an online finding aid include the following elements:

```
<ead>
        <eadheader>
                <eadid> . . . </eadid>
                <filedesc>
                        <titlestmt>
                                <titleproper> . . . </titleproper>
                        </titlestmt>
                </filedesc>
        </eadheader>
        <archdesc level="collection">
                <did>
                        <unittitle> . . . </unittitle>
                </did>
        </archdesc>
</ead>
```

In the above example, "collection" is one of the valid values for the *level* attribute. Also, the Descriptive Identification <**did**> element requires a sub-element, of which Title of the Unit <**unittitle**> is one option. EAD3 was released in 2015 and is currently being adopted by the archival community. For a complete example of an EAD3-encoded document, see Appendix D.

ONIX (Online Information Exchange) DTD and Schema

ONIX[29] is a family of standard XML formats that publishers and others in the book trade can use to distribute *ONIX messages*—electronic information about their books, serials, and other publications. ONIX is overseen by EDItEUR, "the international group coordinating development of the standards infrastructure for electronic commerce in the book, e-book and serials sectors."[30] In the online environment, ONIX messages replace the information that would be found if one were looking at a jacket cover or a container (and often contain even more metadata about a resource). ONIX addresses both the metadata structure (the elements of which are described in Chapter 7) and the encoding of the metadata. XML was chosen as the encoding format because it is optimized for data exchange between computers, the tags are human readable, and XML software is inexpensive enough even for small publishers. In the publishing industry, *ONIX for Books* is the most widely used metadata standard for describing print and electronic books. The latest version of ONIX for Books is Release 3 revision 0.3. It comprises five parts:

- **ONIX for Books 3.0.3 Product Information Format Specification—**a comprehensive guide to the complete ONIX for Books format. It includes overviews of the ONIX product record, the XML data elements and tags, the top level XML message structure, and the message header; it also provides a complete sample ONIX for Books message.
- **ONIX for Books 3.0 Acknowledgment Specification—**a format for sending response messages to confirm receipt of ONIX for Books data.

- **Implementation and Best Practice Guide**—a content standard that guides users to minimize variation in ONIX for Books data.
- **Schema Definitions**—ONIX for Books may be encoded using three different XML structures: a DTD, a schema, and using RELAX NG (RNG; another XML schema language).[31] Although a DTD is available, it is not recommended; the XML schema and the RNG are the preferred methods for structuring ONIX messages.
- **Code lists**—ONIX code lists are available in a variety of forms. These lists control the values entered into certain ONIX elements—these are forms of controlled vocabulary.[32]

All of these components are found on the EDItEUR website. For an example showing an excerpt from an ONIX for Books message, see Figure 6.9.

```xml
<ONIXMessage release="3.0">
  <Header>
        <Sender>
                <SenderName>ABC-CLIO</SenderName>
                <ContactName>Jane Doe</ContactName>
                <EmailAddress>Jane_Doe@abc-clio.com</EmailAddress>
        </Sender>
  </Header>
  <Product>
        <RecordReference>9781598848571</RecordReference>
        <NotificationType>03</NotificationType>
        <ProductIdentifier>
                <ProductIDType>15</ProductIDType>
                <IDValue>9781598848571</IDValue>
        </ProductIdentifier>
        <Barcode>
                <BarcodeType>01</BarcodeType>
                <PositionOnProduct>00</PositionOnProduct>
        </Barcode>
        <DescriptiveDetail>
                <ProductComposition>00</ProductComposition>
                <ProductForm>BB</ProductForm>
                <NoCollection/>
                <TitleDetail>
                        <TitleType>01</TitleType>
                        <TitleElement>
                                <TitleElementLevel>01</TitleElementLevel>
                                <TitleText>Introduction to Cataloging and Classification</TitleText>
                        </TitleElement>
                </TitleDetail>
                <Contributor>
                        <SequenceNumber>1</SequenceNumber>
                        <ContributorRole>A01</ContributorRole>
                        <PersonName>Daniel N. Joudrey</PersonName>
                        <PersonNameInverted>Joudrey, Daniel</PersonNameInverted>
                </Contributor>
                <Contributor>
                        <SequenceNumber>2</SequenceNumber>
                        <ContributorRole>A01</ContributorRole>
                        <PersonName>Arlene G. Taylor</PersonName>
                        <PersonNameInverted>Taylor, Arlene</PersonNameInverted>
                </Contributor>
                <EditionNumber>11</EditionNumber>
                <Extent>
                        <ExtentType>00</ExtentType>
                        <ExtentValue>1048</ExtentValue>
                        <ExtentUnit>03</ExtentUnit>
                </Extent>
```

```xml
            <Subject>
                <MainSubject/>
                    <SubjectSchemeIdentifier>10</SubjectSchemeIdentifier>
                    <SubjectCode>LAN025030</SubjectCode>
                    <SubjectHeadingText>LANGUAGE ARTS & DISCIPLINES / Library &
                    Information Science / Cataloging & Classification</SubjectHeadingText>
            </Subject>
        </DescriptiveDetail>
        <CollateralDetail>
            <TextContent>
                <TextType>03</TextType>
                <ContentAudience>00</ContentAudience>
                <Text textformat="02"> A new edition of this best-selling textbook reintroduces
                the topic of library cataloging from a fresh, modern perspective. <br/> • Delineates
                the new cataloging landscape• Shares a principles-based perspective • Provides
                introductory text for beginners and intermediate students • Emphasizes descriptive
                and subject cataloging, as well as format-neutral cataloging • Covers new cataloging
                rules and RDA</Text>
            </TextContent>
            <TextContent>
                <TextType>03</TextType>
                <ContentAudience>00</ContentAudience>
                <Text>Not many books merit an eleventh edition, but this popular text does. Newly
                updated, <i>Introduction to Cataloging and Classification</i> provides an
                introduction to descriptive cataloging based on contemporary standards, explaining
                the basic tenets to readers without previous experience, as well as to those who
                merely want a better understanding of the process as it exists today. The text opens
                with the foundations of cataloging, then moves to specific details and subject matter
                such as Functional Requirements for Bibliographic Records (FRBR),…. </Text>
            </TextContent>
            <TextContent>
                <TextType>12</TextType>
                <ContentAudience>00</ContentAudience>
                <Text textformat="02"><b>Daniel N. Joudrey</b>, MLIS, PhD, is associate
                professor, School of Library and Information Science, Simmons College, Boston,
                MA. <b>Arlene G. Taylor</b>, MSLS, PhD, is professor emerita, School of
                Information Sciences, University of Pittsburgh, ….</Text>
            </TextContent>
            <TextContent>
                <TextType>06</TextType>
                <ContentAudience>00</ContentAudience>
                <Text textformat="02">"I recommend this book to instructors and students, to
                practicing professionals and paraprofessionals, and to selectors for libraries that
                support library science curricula. Ideally, this text would be used in conjunction with
                a wide variety of practical exercises in original cataloging and the creation of
                authority records." - <strong>Technical Services Quarterly</strong></Text>
            </TextContent>
        </CollateralDetail>
                                  . . . [MUCH MORE DETAIL] . . .
    </Product>
</ONIXMessage>
```

Figure 6.9. An Excerpt from an ONIX for Books Message. (ONIX coding shown in bold)

MARCXML Schema and MODS

In the mid-1990s LC's Network Development and MARC Standards Office (NDMSO) developed two SGML DTDs to provide access to MARC data in an SGML format without a loss of data.[33] As technology evolved, these two SGML DTDs were converted to XML DTDs. The two MARC XML DTDs defined all the elements that appear in a MARC record and specified how they were tagged and represented with XML coding. But as DTDs have gone out of fashion in favor of schemas, the MARC DTDs have been retired.

Two XML schemas have been developed by NDMSO: a MARCXML Schema[34] and MODS (Metadata Object Description Schema).[35] The MARCXML Schema supports XML markup of full MARC 21 records, featuring lossless conversion to and from MARC 21 records, a conversion toolkit, and style sheets. It simply duplicates the MARC content designation structure in a native XML encoding format. For an example of a MARCXML-encoded record, see Figure 6.10.

MODS is an XML schema that is intended to be able to carry selected data from existing MARC 21 records in an XML environment. It is also used to create original metadata records. It includes a subset of MARC fields and uses language-based tags rather than numeric ones (e.g., **<title>** is used rather than 245). In some cases there has been a regrouping of elements from the MARC 21 bibliographic format. Because it is only a subset of MARC, after records have been converted to MODS, the MODS records cannot be converted back to MARC 21 records without a loss of data. Version 3.6 of MODS was made available in 2015. For an example of a MODS bibliographic record, see Figure 7.5 on page 339–40. For information about descriptive elements used in TEI Headers, EAD, ONIX, and MODS, see Chapter 7.

BIBFRAME: A FUTURE STANDARD?

At the time of this writing, a new approach to encoding is under development at LC. It is intended to move library metadata from the record-bound MARC structure to RDF-based individual metadata statements that are not confined to a catalog. In other words, the approach in development embraces the linked data format that is the foundation of the Semantic Web (discussed in Chapter 1). The LC MARC-replacement project is called the Bibliographic Framework Initiative but is more commonly referred to as *BIBFRAME*.

```
<record xmlns="http://www.loc.gov/MARC21/slim"
xmlns:cinclude="http://apache.org/cocoon/include/1.0" xmlns:zs="http://www.loc.gov/zing/srw/">
        <leader>02384cam a2200457 i 4500</leader>
        <controlfield tag="001">18660797</controlfield>
        <controlfield tag="005">20160818094544.0</controlfield>
        <controlfield tag="008">150616s2015 cau b 001 0 eng</controlfield>
        <datafield tag="010" ind1=" " ind2=" ">
                <subfield code="a">2015012911</subfield></datafield>
        <datafield tag="020" ind1=" " ind2=" ">
                <subfield code="a">9781598848571 (hardback : acid-free paper)</subfield></datafield>
        <datafield tag="020" ind1=" " ind2=" ">
                <subfield code="a">9781598848564 (pbk : acid-free paper)</subfield></datafield>
        <datafield tag="020" ind1=" " ind2=" ">
                <subfield code="z">9781440837456 (ebook)</subfield></datafield>
        <datafield tag="040" ind1=" " ind2=" ">
                <subfield code="a">DLC</subfield>
                <subfield code="b">eng</subfield>
                <subfield code="c">DLC</subfield>
                <subfield code="e">rda</subfield>
                <subfield code="d">DLC</subfield></datafield>
        <datafield tag="050" ind1="0" ind2="0">
                <subfield code="a">Z693</subfield>
                <subfield code="b">.W94 2015</subfield></datafield>
        <datafield tag="082" ind1="0" ind2="0">
                <subfield code="a">025.3</subfield>
                <subfield code="2">23</subfield></datafield>
        <datafield tag="084" ind1=" " ind2=" ">
                <subfield code="a">LAN025030</subfield>
                <subfield code="a">LAN020000</subfield>
                <subfield code="2">bisacsh</subfield></datafield>
        <datafield tag="100" ind1="1" ind2=" ">
                <subfield code="a">Joudrey, Daniel N.,</subfield>
                <subfield code="e">author.</subfield></datafield>
        <datafield tag="245" ind1="1" ind2="0">
                <subfield code="a">Introduction to cataloging and classification /</subfield>
                <subfield code="c">Daniel N. Joudrey, Arlene G. Taylor, and David P. Miller.</subfield>
        </datafield>
        <datafield tag="250" ind1=" " ind2=" ">
                <subfield code="a">Eleventh edition.</subfield></datafield>
        <datafield tag="264" ind1=" " ind2="1">
                <subfield code="a">Santa Barbara, California :</subfield>
                <subfield code="b">Libraries Unlimited,</subfield>
                <subfield code="c">[2015]</subfield></datafield>
        <datafield tag="300" ind1=" " ind2=" ">
                <subfield code="a">xxv, 1048 pages :</subfield>
                <subfield code="b">illustrations ;</subfield>
                <subfield code="c">27 cm.</subfield></datafield>
        <datafield tag="336" ind1=" " ind2=" ">
                <subfield code="a">text</subfield>
                <subfield code="2">rdacontent</subfield></datafield>
```

```
        <datafield tag="337" ind1=" " ind2=" ">
                <subfield code="a">unmediated</subfield>
                <subfield code="2">rdamedia</subfield>
        </datafield>
        <datafield tag="338" ind1=" " ind2=" ">
                <subfield code="a">volume</subfield>
                <subfield code="2">rdacarrier</subfield>
        </datafield>
        <datafield tag="490" ind1="0" ind2=" ">
                <subfield code="a">Library and information science text series</subfield></datafield>
        <datafield tag="500" ind1=" " ind2=" ">
                <subfield code="a">Significant expansion of: Introduction to cataloging and classification.
                Tenth edition / Arlene G. Taylor ; with the assistance of David P. Miller.</subfield>
        </datafield>
        <datafield tag="504" ind1=" " ind2=" ">
                <subfield code="a">Includes bibliographical references (pages 1001-1022) and index.
                </subfield></datafield>
        <datafield tag="650" ind1=" " ind2="0">
                <subfield code="a">Descriptive cataloging.</subfield></datafield>
        <datafield tag="650" ind1=" " ind2="0">
                <subfield code="a">Subject cataloging.</subfield></datafield>
        <datafield tag="650" ind1=" " ind2="0">
                <subfield code="a">Classification</subfield>
                <subfield code="x">Books.</subfield>
        </datafield>
        <datafield tag="630" ind1="0" ind2="0">
                <subfield code="a">Resource description & access.</subfield></datafield>
        <datafield tag="650" ind1=" " ind2="7">
                <subfield code="a">LANGUAGE ARTS & DISCIPLINES / Library & Information Science /
                Cataloging & Classification.</subfield>
                <subfield code="2">bisacsh</subfield>
        </datafield>
        <datafield tag="700" ind1="1" ind2=" ">
                <subfield code="a">Taylor, Arlene G.,</subfield>
                <subfield code="d">1941-</subfield>
                <subfield code="e">author.</subfield>
        </datafield>
        <datafield tag="700" ind1="1" ind2=" ">
                <subfield code="a">Miller, David P.</subfield>
                <subfield code="q">(David Peter),</subfield>
                <subfield code="d">1955-</subfield>
                <subfield code="e">author.</subfield>
        </datafield>
        <datafield tag="700" ind1="1" ind2=" ">
                <subfield code="a">Taylor, Arlene G.,</subfield>
                <subfield code="d">1941-</subfield>
                <subfield code="t">Introduction to cataloging and classification.</subfield>
                <subfield code="i">Based on (work):</subfield>
        </datafield>
</record>
```

Figure 6.10. An Excerpt from a MARCXML Record. (MARCXML coding shown in bold)

In the early twenty-first century there were increasingly urgent calls for the replacement of MARC as an encoding standard for cataloging data. Although MARC was a groundbreaking achievement and has adapted to libraries' needs over time, critics have declared that MARC has reached the end of its useful life; among these is Roy Tennant's provocative and influential article, "MARC Must Die."[36] As the *Functional Requirements for Bibliographic Records* (FRBR) became the underlying model for bibliographic metadata, with its emphasis on identifying multiple and diverse relationships, the MARC format's limitations and inflexibility became clearer. For more on the rationale for moving beyond MARC, please see Joudrey, Taylor, and Miller's *Introduction to Cataloging and Classification*, 11th edition.[37]

The BIBFRAME initiative is an ambitious attempt to replace MARC as the encoding standard for library metadata. At the same time, it intends to retain the huge investment made by libraries in the creation of authoritative and reliable metadata. As the November 2012 LC report, *Bibliographic Framework as a Web of Data*, states:

> Libraries generate, maintain and curate an enormous amount of high-quality data . . . that is valuable well beyond traditional library boundaries. . . . In modeling the MARC 21 format as a Web of Data it is important to deconstruct and then reconstruct the informational assets that comprise MARC.[38]

Beyond retaining the metadata content of the MARC records, BIBFRAME is also designed to continue the basic MARC functions of "representation and communication of bibliographic and related information in machine-readable form."[39] Unlike previous calls for the "death" of MARC, BIBFRAME does not consign the billions of existing MARC records to a sealed-off zone of legacy records but moves their value forward into the Semantic Web through an RDF-based linked data structure (described in Chapter 5). BIBFRAME, therefore, relies on RDF triples (subject–predicate–object) for modeling its metadata statements.

BIBFRAME 2.0, the current iteration of the framework, comprises two major components: the *BIBFRAME Model* and the *BIBFRAME Vocabulary*. The BIBFRAME Vocabulary is "the key to the description of resources."[40] Similar to the defined set of tags and subfields for the purpose of content designation found in the MARC bibliographic format, the BIBFRAME Vocabulary has a defined set of *classes* (groups of things or entities) and *properties* (relationships or attributes) to describe bibliographic resources. The BIBFRAME Vocabulary, at this writing, comprises over 60 classes or subclasses (a brief sample of which is provided in Table 6.2). There are also well over 100 properties identified; some of which may be seen in Table 6.3.

Table 6.2. A Small Sample of BIBFRAME Classes and Associated Subclasses.

Classes	Associated Subclasses
AdminMetadata	**DescriptionAuthentication** **DescriptionConventions** **GenerationProcess**
Agent	**Family** **Jurisdiction** **Meeting** **Organization** **Person**
Carrier	**n/a**
Classification	**ClassificationDdc** **ClassificationLcc** **ClassificationNlm** **ClassificationUdc**
DigitalCharacteristic	**CartographicDataType** **CartographicObjectType** **EncodedBitrate** **EncodingFormat** **FileSize** **FileType** **ObjectCount** **RegionalEncoding** **Resolution**
Instance	**Archival** **Electronic** **Manuscript** **Print** **Tactile**
Title	**VariantTitle**
VariantTitle	**AbbreviatedTitle** **CollectiveTitle** **KeyTitle** **ParallelTitle**

Work	Audio Cartography Dataset MixedMaterial MovingImage Multimedia NotatedMovement NotatedMusic Object StillImage Text

Table 6.3. A Small Sample of BIBFRAME Properties.

Properties	Associated Class
dimensions	Instance
dissertation	Work or Instance
duration	Work or Instance
edition	Classification
editionEnumeration	Instance
editionStatement	Instance
electronicLocator	Item
emulsion	Instance
ensemble	Work
ensembleType	MusicEnsemble
enumerationAndChronology	Item
equinox	Cartographic
eventContent	Event
eventContentOf	Work
exclusionGRing	Cartographic
expressionOf	Work
extent	Instance
findingAid	Work or Instance

The BIBFRAME Model describes the structure of bibliographic resources and their potential relationships to other entities; it comprises three main high-level entities: *work*, *instance*, and *item*. This structure might seem somewhat familiar; these entities are similar to FRBR's Group 1 entities (discussed in Chapter 5). *Work* in the BIBFRAME Model is the "conceptual essence of the cataloged resource";[41] it incorporates both the work and expression levels from FRBR. *Instance*—one or more material embodiments of a work—is very much equivalent to FRBR's manifestation; the definition of *item*, too, is very similar in both standards. The BIBFRAME Model's item is "an actual copy (physical or electronic) of an Instance."[42] In addition to the three high-level entities, the BIBFRAME Model identifies three core classes: *agents*, *subjects*, and *events*. *Agents* are entities associated with a work or an instance, such as persons, corporate bodies, families, governments, and so on. Agents play roles in the creation of the work or the instance. An agent could be an author, a publisher, an illustrator, a translator, and the like. *Subjects* reflect what the resource is about, and *events* are occurrences that might be related to the content of a work (e.g., the Battle of Gettysburg).[43] Additional classes may also be used in the model to describe relationships and attributes of resources. The very basic conceptual model of BIBFRAME 2.0 is illustrated in Figure 6.11; an example is provided next to the model. An encoded BIBFRAME description for the item in Figure 6.4 is presented in Appendix E because the figure is too large to include in this chapter. Whereas the MARC record for the same item fills a page (or a little over one page), the equivalent data in BIBFRAME 2.0 is approximately 10 times larger.

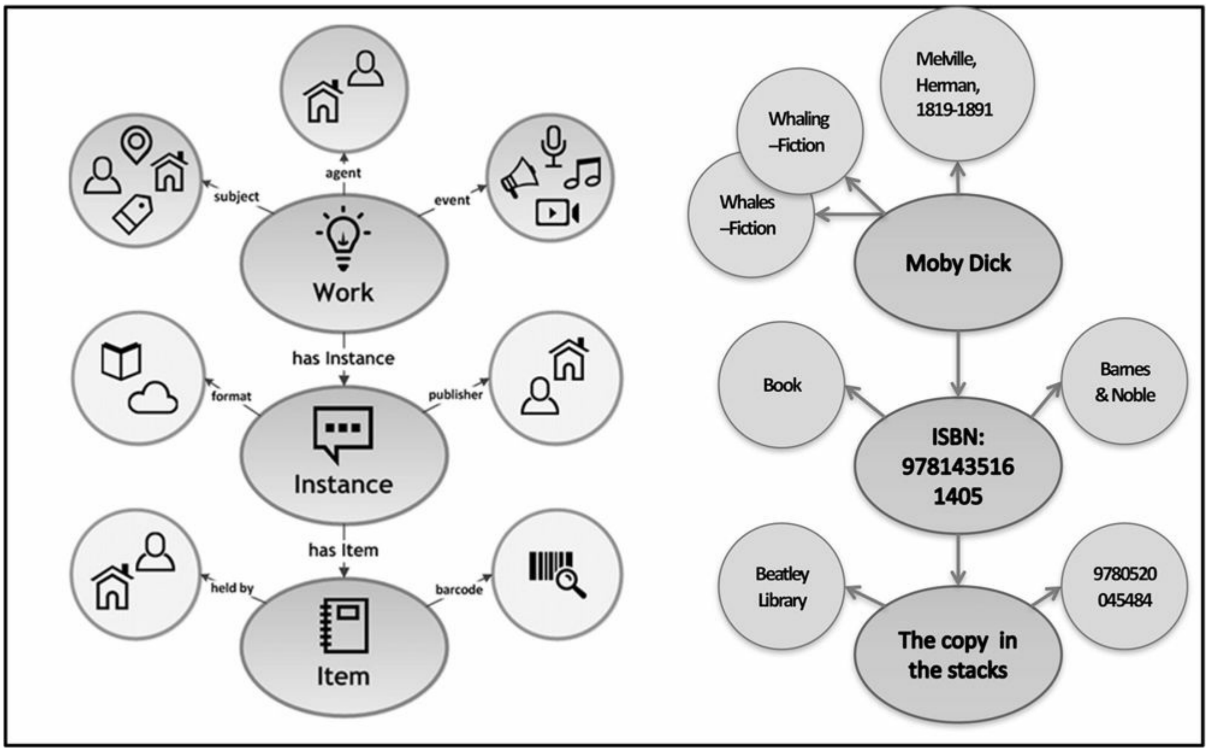


Figure 6.11. The Basic Conceptual Model for BIBFRAME 2.0 with an Example to Illustrate the Concepts. (Source of Model: Library of Congress, BIBFRAME 2.0 Model, https://www.loc.gov/bibframe/docs/bibframe2-model.html)

CONCLUSION

This chapter has discussed some of the ways of encoding metadata. We have reviewed the MARC format that has long been used in library catalogs. Although it has in many ways done most everything it has been asked to do, there are difficulties associated with this aging standard. Over more recent decades many have suggested that an SGML- or XML-based approach to encoding metadata was the answer. As technology continues to develop, however, the creation of natural language tagged XML records seems somewhat out of date. Progress continues and another approach is offered: using RDF-structured linked data statements to record and share descriptive metadata. This approach is not yet ready for consumption, but there are many LIS professionals working toward developing this new approach.

An understanding of encoding has been of great importance over the past 40 years. Catalogers, archivists, and other information professionals have been responsible not only for the descriptive metadata content but also for the encoding process meant to share that data. It was necessary to understand the containers that hold the metadata and how to manipulate the encoding as needed. As technology continues to develop, though, there may be less need to be actively involved in the direct encoding of the information. We may come to rely more on automated template tools or editors that will take our descriptive content in, convert that information to RDF triples behind the scenes, and then present metadata that is ready to share in the linked data environment. In short, while catalogers truly needed to understand MARC in order to catalog, that level of interaction may not be necessary if we adopt BIBFRAME.

In the next chapter, standards for the creation of metadata are reviewed. Included are discussions of content standards, structure standards, and presentation standards.

Some Important Terms in This Chapter

Attribute	Element	ONIX Message
BIBFRAME	Encoding	Romanization
Codes	Entity	Statement of Responsibility
Cooperative Cataloging	Field	Subfield
Data Element Identifier	Fixed Field	Subfield Code
Delimiter	Indicator	Syntax
Directory	Leader	Tag
Document Type Definition	Linked Data	Title Proper
	Meta-language	XML Schema

Some Important Acronyms in This Chapter

ANSI: American National Standards Institute
ASCII: American Standard Code for Information Interchange
BIBFRAME: Bibliographic Framework Initiation
DTD: Document Type Definition
EAD: Encoded Archival Description
EBCDIC: Extended Binary Coded Decimal Interchange Code
GML: Generalized Markup Language
HTML: Hypertext Markup Language
IBM: International Business Machines
ILS: Integrated Library System
ISO: International Organization for Standardization
LC-MARC: Library of Congress MARC
MARC: MAchine-Readable Cataloging
MODS: Metadata Object Description Schema
NDMSO: Library of Congress Network Development and MARC Standards Office
NISO: National Information Standards Organization
ONIX: Online Information Exchange
RDF: Resource Description Framework
RDFa: RDF in Attributes
RNG: Relax (Regular Language for XML) Next Generation
SGML: Standard Generalized Markup Language
TEI: Text Encoding Initiative
UCS: *Universal Coded Character Set*
UNIMARC: Universal MARC
URI: Uniform Resource Identifier
USMARC: United States MARC
W3C: World Wide Web Consortium
XHTML: Extensible Hypertext Markup Language
XML: Extensible Markup Language

NOTES

All URLs accessed June 2017.

1. Henriette Avram, "Machine Readable Cataloging (MARC): 1961–1974 [Classic from 1975]," in *Encyclopedia of Library and Information Science*, 3rd ed., eds. Marcia J. Bates and Mary Niles Maack (New

York: Taylor and Francis, 2009), 3520.

2. Kevin S. Clarke, "Extensible Markup Language (XML)," in *Encyclopedia of Library and Information Science*, 3rd ed., eds. Marcia J. Bates and Mary Niles Maack (New York: Taylor and Francis, 2009), 1808.

3. Ibid., 1809.

4. Unicode Consortium, "Unicode 10.0.0, 2017 June Announcement," http://www.unicode.org/versions/Unicode10.0.0/.

5. Karen Smith-Yoshimura, Catherine Argus, Timothy J. Dickey, Chew Chiat Naun, Lisa Rowlinson de Ortiz, and Hugh Taylor, *Implications of MARC Tag Usage on Library Metadata Practices*, a report produced by OCLC Research in support of the RLG Partnership. (Dublin, OH: OCLC, 2010), 19–20, http://www.oclc.org/research/publications/library/2010/2010-06.pdf.

6. William E. Moen and Penelope Benardino, "Assessing Metadata Utilization: An Analysis of MARC Content Designation Use," *International Conference on Dublin Core and Metadata Applications* (S.l.: DCMI, 2003), 171–80, http://dcpapers.dublincore.org/pubs/article/view/745.

7. This is a holdover from card catalog days when an additional card for the title proper was not made when the title was already the main entry (i.e., the primary access point because no creator was identified or too many [more than three] creators were listed). In most cases now, it functions as an indicator of whether a title is the primary access point for the record (0 = yes; 1= no).

8. National Information Standards Organization, *ANSI/NISO Standard Z39.2-1994 (R2009): Information Interchange Format* (Baltimore, MD: NISO, 1994).

9. International Organization for Standardization, *Information and Documentation—Format for Information Exchange*, ISO 2709:2008, 4th ed. (Geneva: ISO, 2008).

10. Library of Congress, Network Development and MARC Standards Office, "MARC 21 Specifications for Record Structure, Character Sets, and Exchange Media: Character Sets and Encoding Options," http://www.loc.gov/marc/specifications/speccharintro.html.

11. Daniel N. Joudrey, Arlene G. Taylor, and David P. Miller, *Introduction to Cataloging and Classification*, 11th ed. (Santa Barbara, CA: Libraries Unlimited, 2015), Chapter 21.

12. IFLA, *UNIMARC Manual: Bibliographic Format*, 3rd ed., ed. Alan Hopkinson (Munich: K.G. Saur, 2008).

13. Airi Salminen, "Markup Languages," in *Encyclopedia of Library and Information Science*, 3rd ed., eds. Marcia J. Bates and Mary Niles Maack (New York: Taylor and Francis, 2009), 3563.

14. International Organization for Standardization, *Information Processing: Text and Office Systems: Standard Generalized Markup Language (SGML)*, ISO 8879:1986 (Geneva: ISO,1988).

15. W3C, "HTML," http://www.w3.org/html/.

16. W3C, "XHTML 1.0: The Extensible HyperText Markup Language (Second Edition): A Reformulation of HTML 4 in XML 1.0. W3C Recommendation," revised 2002, http://www.w3.org/TR/xhtml1/.

17. W3C, "HTML5," 2014, https://www.w3.org/TR/html5/.

18. Kristine Schachinger, "How to Use HTML Meta Tags," *Search Engine Watch*, 2001, https://searchenginewatch.com/sew/how-to/2067564/how-to-use-html-meta-tags.

19. W3C, "HTML Microdata," 2017, https://www.w3.org/TR/microdata/.

20. "Microformats," http://microformats.org/.

21. W3C, "RDFa 1.1 Primer: Rich Structured Data Markup for Web Documents," 3rd ed., 2015, https://www.w3.org/TR/rdfa-primer/.

22. W3C, "XML in 10 Points," https://www.w3.org/XML/1999/XML-in-10-points-19990327.

23. W3C, "XML Schema," http://www.w3schools.com/xml/xml_schema.asp.

24. Text Encoding Initiative, "TEI Guidelines," http://www.tei-c.org/Guidelines/.

25. Text Encoding Initiative, "TEI: History," http://www.tei-c.org/About/history.xml.

26. Text Encoding Initiative, "TEI Lite," http://www.tei-c.org/Guidelines/Customization/Lite/.

27. Text Encoding Initiative, "TEI: Text Encoding Initiative," http://www.tei-c.org/index.xml.

28. "EAD: Encoded Archival Description," Library of Congress, http://www.loc.gov/ead/.

29. EDItEUR, "ONIX," http://www.editeur.org/8/ONIX/.

30. EDItEUR, "About," http://www.editeur.org/2/About/.

31. RELAX NG stands for REgular LAnguage for XML Next Generation.

32. EDItEUR, "Release 3.0 Downloads," http://www.editeur.org/93/Release-3.0-Downloads/.

33. Library of Congress, Network Development and MARC Standards Office, "MARCXML: MARC 21 XML Schema," http://www.loc.gov/standards/marcxml/.

34. Ibid.

35. Library of Congress, Network Development and MARC Standards Office, "MODS: Metadata Object Description Schema," http://www.loc.gov/standards/mods/.

36. Roy Tennant, "MARC Must Die," *Library Journal* 127, no. 17 (2002): 26–27.

37. Joudrey, Taylor, and Miller, *Introduction to Cataloging and Classification*, 11th ed., Chapter 22.

38. Library of Congress, "Bibliographic Framework as a Web of Data: Linked Data Model and Supporting Services," http://www.loc.gov/bibframe/pdf/marcld-report-11-21-2012.pdf.

39. Kevin M. Ford, "LC's Bibliographic Framework Initiative and the Attractiveness of Linked Data," *Information Standards Quarterly* 24, no. 2/3 (2012): 47.

40. Library of Congress, "BIBFRAME Frequently Asked Questions," http://www.loc.gov/bibframe/faqs/.

41. Library of Congress, "Overview of the BIBFRAME 2.0 Model," http://www.loc.gov/bibframe/docs/bibframe2-model.html.

42. Ibid.

43. Ibid.

SUGGESTED READINGS

"EAD: Encoded Archival Description." Library of Congress. http://www.loc.gov/ead/.

EDItEUR. "ONIX: Overview." http://www.editeur.org/83/Overview/.

Furrie, Betty. *Understanding MARC Bibliographic: Machine-Readable Cataloging.* 8th ed. Washington, DC: Library of Congress, Cataloging Distribution Service, 2009. http://www.loc.gov/marc/umb.

Godby, Carol Jean. *Mapping ONIX to MARC*. Dublin, OH: OCLC Research, 2010.

Joudrey, Daniel N., Arlene G. Taylor, and David P. Miller. *Introduction to Cataloging and Classification*. 11th ed. Santa Barbara, CA: Libraries Unlimited, 2015. Chapter 21: MARC Encoding.

Library of Congress, Network Development and MARC Standards Office. "BIBFRAME: Bibliographic Framework Initiative." http://www.loc.gov/bibframe/.

Library of Congress, Network Development and MARC Standards Office. "MARC Standards." http://www.loc.gov/marc/.

Library of Congress, Network Development and MARC Standards Office. "MARCXML: MARC 21 XML Schema." http://www.loc.gov/standards/marcxml/.

Library of Congress, Network Development and MARC Standards Office. "MODS: Metadata Object Description Schema." http://www.loc.gov/standards/mods/.

Text Encoding Initiative. "TEI: P5 Guidelines." Version 3.1.0. Last updated on December 15, 2016. http://www.tei-c.org/release/doc/tei-p5-doc/en/html/index.html.

W3C. "Extensible Markup Language." http://www.w3.org/XML/.

W3C. "HTML." http://www.w3.org/html/.

W3C. "HTML5 Tutorial." http://www.w3schools.com/html/.

W3C. "XML Tutorial." http://www.w3schools.com/xml/.

CHAPTER 7

RESOURCE DESCRIPTION

Loosely speaking, there are three parts to creating metadata for an information resource: (1) providing a description of the resource along with other information necessary for the management, preservation, and structure of the resource, (2) providing for access to this description, and (3) providing the syntax of the metadata (i.e., encoding). Chapter 5 gives a general introduction to metadata, and encoding is discussed in Chapter 6. This chapter discusses resource description, and Chapters 8–11 discuss various types of access to metadata and how such access may be provided.

In libraries, *bibliographic record* is the name that has been applied to the description of tangible information resources (e.g., books, sound recordings, maps) for many years. Even though it has been applied to records created for motion pictures, sound recordings, electronic resources, and the like, the word *bibliographic* at various points has had a stigma attached to it arising from the use of *biblio-*, meaning *book*. At times, the term *surrogate record* has been used instead. A surrogate stands in place of someone or something else. The term can be used for a record representing any kind of resource in any kind of information retrieval system. More recently, as the longstanding focus on the record structure has lessened due to our explorations into linked data, some have begun to refer to our bibliographic data simply as *descriptions*, *metadata*, *descriptive metadata*, *resource descriptions,* or *metadata statements*. All of these terms may be used somewhat interchangeably in this and the following chapters.

Some definitions are in order before discussing the creation of resource descriptions. An *information resource* is an instance of recorded information (e.g., a book, article, streaming video, sound recording, online journal). *Descriptive data* is data derived from a resource and used to describe it; it is a representation of the most important characteristics (i.e., *attributes*) of a resource that can help users with activities collectively referred to as *user tasks* (i.e., to find, identify, select, obtain, and explore resources in a collection). The attributes include both descriptive data (such as title proper, dates, extent, resource types, standard number, etc.) and access points (such as creator and contributor names, subjects, work titles, titles of related works, etc.). A *metadata statement* is a standalone description of a single attribute of a resource. For example:

<dc:title>Cataloging and Acquisitions Home**</dc:title>**

A *surrogate record* is a somewhat self-contained collection of metadata statements that describe the characteristics of an information resource. Such a record stands in place of (i.e., is a surrogate for) a resource in retrieval systems such as catalogs, indexes, and bibliographies, while metadata statements may be encoded in documents directly or kept in *triplestores* on the web. Both metadata statements and surrogate records contain descriptive metadata. In metadata records or statements, a particular piece of descriptive data may be referred to as the *content* or the *value* assigned to an *element.* An *element* is a field in a metadata schema that represents an important *attribute*. An *access point* is a term (word, heading, prescribed string, etc.) in a surrogate record that is used to retrieve that record. Access points are often singled out from other descriptive data and are placed under authority control (see discussion in Chapter 8).

A file of surrogate records serves as a filter to keep a user from having to search through myriad irrelevant resources. The descriptions must be distinctive enough that they cannot be confused with descriptions in records for any other resources. A surrogate record's most important function is to assist the user in evaluating the possibility that the resource it represents will be useful and will contain information that the user wishes to explore further. Resource descriptions are most helpful when they are predictable in both form and content. Adherence to standards ensures such predictability. At the very least, most standards dictate which metadata elements are required and which are optional. Some are quite extensive in their guidance, while others provide only minimal instruction. Some existing standards are discussed below, but metadata standards are constantly being revised, updated, and re-imagined. It is, therefore, vitally important that readers consult official documentation for each standard to find the most current information.

SOME PRELIMINARY CONSIDERATIONS

The process of description, no matter the environment in which it is undertaken, begins with an examination of the resource to be described. This examination helps organizers to get their bearings and gain an understanding of the tasks ahead. Getting a sense of the scope, size, form, and contents of a resource is good preparation for describing it. Daniel N. Joudrey, Arlene G. Taylor, and David P. Miller state that the description process begins with a series of decisions or considerations.[1] For example, in libraries, a cataloger might begin by asking some of the following questions:

- What type of resource is this?
- How was it issued?
- Is the whole resource, or just a part of it, being described?
- From which sources should the metadata be taken?
- Is this a new resource?
- Is it related to other resources?
- What characteristics are important to describe?

It is necessary to determine exactly what is to be described. In *Cataloging Cultural Objects* (CCO)—a content standard for art and cultural heritage collections—the general guidelines encourage organizers to think about the resource systematically: "To catalog a work is to describe what it is, who made it, where it was made, how it was made, the materials of which it was made, and what it is about."[2] This directive is applicable to all information resources in all information environments. The organizer must always start with a simple question: "What am I describing?"

Resource Types

Until recently, catalogers began the description process by determining the *resource type*. A resource type is a category that reflects the nature or overall form of the resource (i.e., whether it is a book, a map, a serial, etc.). This designation determined the instructions to be followed when creating metadata for that resource (e.g., there were separate rules for maps and other nonbook formats). The emergence of electronic resources, to some extent, threw many traditional resource-type designations into chaos. For example, it was no longer quite as straightforward to describe an online map published in parts over periods of time. Catalogers were not always sure which instructions to follow when a resource fell outside the parameters of a single well-defined resource type.

Although organizers are *still* concerned with identifying the type of resource in hand, it is no longer the primary determining factor in how something is described. The current library cataloging content standard, *RDA: Resource Description & Access*,[3] changed the focus so that the basis of description is no longer centered on particular resource types. Instead, RDA simplifies the situation by looking at common characteristics across all resource types and by organizing the description around the different components of the International Federation of Library Associations and Institutions' (IFLA) *Functional Requirements for Bibliographic Records* (FRBR)[4] and *Library Reference Model* (LRM).[5] RDA contains instructions for describing a resource, regardless of whether it is a sound recording, a manuscript, a three-dimensional object, or another type of resource.

In other information environments, resource types also play a role. In cultural heritage collections using CCO, *work type* is an equivalent concept to *resource type*. "Work Type typically refers to a work's physical form, function, or medium (for example, sculpture, altarpiece, cathedral, storage jar, painting, etching)."[6] It is one of the required elements in a description. In archives using *Describing Archives: A Content Standard* (DACS), material types may be brought out in the title devised for the collection if one or two specific forms predominate.[7]

FRBR Entities

Another concern, particularly in libraries, is whether organizers should describe a *work* (the intellectual creation) or an *item* (the physical resource). This is not a new concern; it has been around for decades. This issue was a source of debates between two cataloging giants, Seymour Lubetzky and Michael Gorman, before the adoption of *Anglo-American Cataloguing Rules, Second Edition* (AACR2) in 1978. AACR2 and its editor, Gorman, came down firmly on the side of describing the item in hand. Lubetzky, whose work on the Paris Principles influenced the first edition of the *Anglo-American Cataloging Rules* (AACR or AACR1), believed that the work should be the prime focus.

In the 1990s IFLA took on the challenge of identifying whether to describe works, items, or something else. In 1998 it released its FRBR report, which addresses the issue.[8] (Please see Chapter 5 for a more complete review of FRBR concepts.) The FRBR model identifies four entities that are the products of creation and, therefore, may be described in metadata: *work, expression, manifestation,* and *item* (also known as Group 1 entities). FRBR also discusses *agents* (i.e., Group 2 entities) that are responsible for the existence of resources; these include *person* and *corporate body* (with *family* added later). It also identifies Group 3 entities that can be subjects of works—*concept, object, event,* and *place.* In contemporary library cataloging, the entity that is usually chosen as the starting point for cataloging is the *manifestation,* although metadata elements specifically related to the work and expression levels are also included in descriptions, often as access points. In the case of rare books, the physical characteristics of the *item* are described explicitly. In LRM, the update to the FRBR conceptual model, the entities have changed somewhat. Although works, expressions, manifestations, and items are still described, the group structure has been eliminated, as has the set of Group 3 entities. For more on LRM, please see Chapter 5.

In general, FRBR is a library-centric concept. It does not apply particularly well to the archival environment or to art and other cultural heritage materials. The museum community uses its own conceptual model (CIDOC CRM), and the archival community is developing one (RiC), both of which are discussed in Chapter 5.

Mode of Issuance and Level of Description

In order to create a description, organizers must determine the resource's mode of issuance. *Mode of issuance* refers to

- how the resource is issued (e.g., Is the resource published in separate parts or issues? Or, is it distributed as one unit?),
- how the resource is updated (e.g., Are there supplements, and if so, are supplements discrete or integrating?), and
- whether there is an expected conclusion to the resource's publication (e.g., Is it an ongoing publication?).

There are four common modes of issuance identified in libraries today:

- **Single Unit**: One physical resource that is complete unto itself is described as a *single unit*. For example, the textbook *Introduction to Cataloging and Classification* is a single unit. A single unit also may be referred to as a *monograph*.
- **Multipart Monograph**: There are resources that are distributed in volumes or parts. Often, they follow one another in succession and have the same title. If these resources are limited in duration (i.e., they are meant to end), then they are known as *multipart monographs*. For example, volume 1 of *The Works of Shakespeare in Two Volumes* is probably not a describable unit by itself, but volumes 1 and 2 together are (i.e., together they make up a multipart monograph). Or, in some cases, there are resources that come in a larger set with the same overarching title (although each can also have its own title); they also may be described as a unit. For example, *Great Books of the Western World* is a set where each volume has its own, often famous, author and title. It might be cataloged as one resource with multiple volumes (i.e., as a multipart monograph), or each volume might be cataloged separately (i.e., each as a single unit).
- **Serial**: There are also resources that are (1) issued in successive volumes or parts that follow one another in succession, (2) known by the same title, (3) often bearing numbering, and (4) meant to

continue indefinitely. These resources are called *serials*. Serials include resources such as scholarly journals, newspapers, popular magazines, and other periodicals. For example, all of the volumes and issues of the journal *Library Resources & Technical Services* are described together in one record as a serial.

- **Integrating Resource**: Finally, there are some resources that are considered single units, but their content is updated periodically. New material is inserted seamlessly into an existing resource in such a way that the discrete update disappears. These are referred to as *integrating resources* and include loose-leaf publications, databases, and websites.

These modes are illustrated in Figure 7.1.

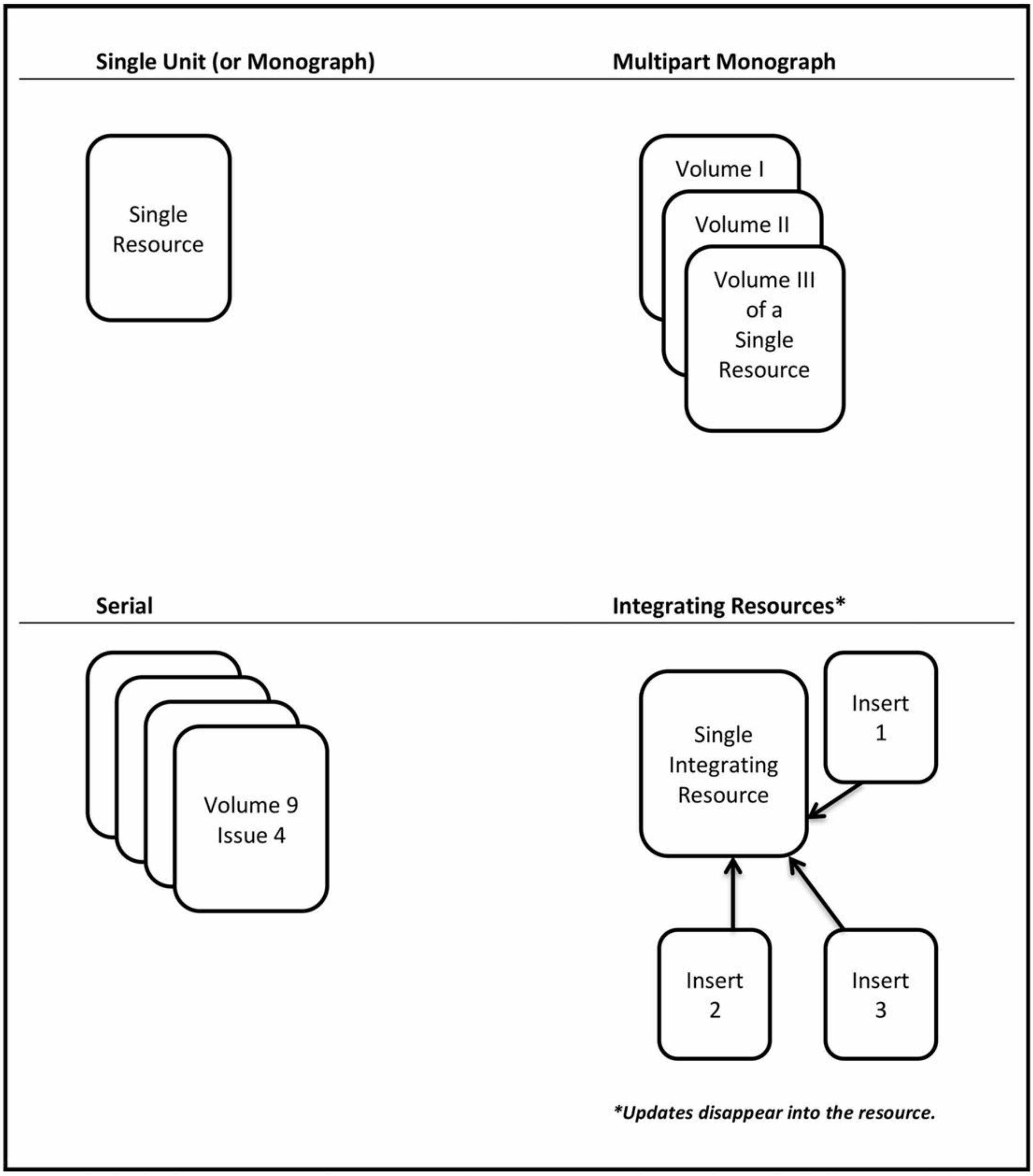


Figure 7.1. Modes of Issuance.

Before RDA introduced the four modes of issuance above, the 2002 update to AACR2 divided resources into two mutually exclusive groups: (1) *finite resources*, those that are complete or have a predetermined conclusion, and (2) *continuing resources*, those that are ongoing (i.e., those which will have additions made to them without a predetermined end in sight). These two categories have also been referred to as *monographs* and *serials,* and they had an organizational impact on technical services divisions for years. The monograph versus serial distinction has been used to set up working departments in many large academic libraries. In some cases, technical services units have been divided so that separate cataloging and acquisitions departments handle monographs, while serials departments handle both the acquisition and cataloging of serials. In other cases, technical services units have been divided into acquisitions and cataloging, with each of those departments further divided into serials and monographs sections. Most integrated library systems (ILSs) have separate modules for general cataloging and for serials management. Most ILS serial modules, however, are designed to handle anything that needs to be received and checked in over a period of time. This can include multiple-volume sets, unnumbered series, and loose-leaf updates, none of which are included in definitions of *serial*.

Once a resource's mode of issuance has been determined, organizers must decide at which level to describe the resource. Typically in libraries, there are three choices for a single resource:

- **Comprehensive Description**: This level entails describing the resource as a whole. The entire resource, not just a part or a portion, is described in the metadata. A single unit, a multipart monograph, a serial, and an integrating resource can be described comprehensively. For example, a 24-volume set of encyclopedias should be described comprehensively; creating separate records for each volume would not be particularly helpful.
- **Analytical Description**: An *analytical description* means that only a part or a portion of the resource is described in the metadata. Again, this can be true of any resource, no matter its mode of issuance. An example is creating a bibliographic record for one short story present in a compilation of nine short stories. As RDA states, "It is possible to prepare separate analytical descriptions for any number of parts of a larger resource (i.e., for one part only, for two or more selected parts, or for all parts of the resource)."[9] In indexing and abstracting services, analytical description is used frequently with journal indexing, where the article is the unit of description.
- **Hierarchical Description**: This entails describing the whole resource (made up of two or more parts) and the individual parts. It combines comprehensive and analytical descriptions. This level is not used in most libraries, but dominates the descriptive approach in archives. The finding aid, the primary tool for description and access of archival materials, is based on a hierarchical structure, in which the collection is described and then its parts.

As mentioned previously, the three levels listed above are focused on the description of a single resource (i.e., manifestation- or item-level description). This, however, is not the only option for description. In many information environments, *collection-level description* may be preferable. Collection-level description entails creating metadata that describes more than one individual resource. This is the approach used in archives, where a collection related to a person, family, or a corporate body is described in the aggregate. In archives, collection-level description is used for both practical and intellectual reasons. On the practical side, many archival collections contain thousands of individual items; it would be very time consuming, expensive, and, at times, counterproductive to create an individual description for each piece of correspondence or each financial record in a collection. More importantly, on the intellectual side, archival documents may be understood and used more effectively in the aggregate, because the value of an archival record is enhanced by the relationships that connect it to the other records within the collection.

Collection-level description is used in libraries also. Sometimes, it is appropriate for special collections of unusual materials that are not traditionally described at the individual resource level (e.g., a collection of theatre programs); sometimes, it may be used for low-priority resources (e.g., a box of donated paperbacks); and at other times, a library may use collection-level description as a temporary measure until full cataloging can be performed. In digital libraries, some collections of resources might be described in the aggregate (e.g., a collection of 500 digital photographs of postal scales), either permanently or temporarily. For some institutions, collection-level description is used to make some basic information about the collection available on the web. Museums, for instance, may provide rudimentary information about their entire collection (or parts thereof) on their websites to ensure some level of retrieval for the public through search engines.[10]

Sources

Another preliminary consideration is determining where to look for the most useful metadata. The preferred sources of information used for library cataloging, based on broad categories of resources, are listed in RDA. Much of the basic descriptive metadata about a resource is transcribed directly from a preferred source of information. If needed information is not found in one of the preferred sources, then alternative locations or features are consulted. The RDA guidelines for determining sources are summarized in Table 7.1.[11]

Archival and museum descriptive standards also articulate valid sources of information. DACS, the content standard used in the United States to guide the process of archival description, states

All the information to be included in archival descriptions must come from an appropriate source, the most common of which is the materials themselves. In contrast to library practice, archivists rarely transcribe descriptive information directly from archival materials; rather, they summarize or interpolate information that appears in the materials or devise information from appropriate external sources, which can include transfer documents and other acquisition records, file plans, and reference works. Each element has one or more prescribed sources of information.[12]

In CCO, a content standard used to guide the process of description in cultural heritage settings, the object being described is the primary source of information about that object. Works of art, however, are not always complete in the information they contain about their origins, provenance, history, and so on. To supplement this, CCO expects catalogers to rely on outside scholarly sources to enhance descriptions, an approach similar to DACS's accompanying documentation. The sources used, when describing a resource according to the CCO guidelines, should be cited clearly in the metadata.[13]

Relationships

As the description process begins, organizers must also determine the relationships that a resource has to other entities, such as other resources, to agents involved in the development of the resource, to subjects discussed, and so on. It is important to identify who or what created, developed, presented, and disseminated the resource. Organizers must ask questions, such as

Table 7.1. Preferred Sources of Information According to *RDA: Resource Description & Access*.

<table>
<tr><th>Category of resource types</th><th>Examples</th><th>Preferred sources of information</th><th>If those are unavailable, then . . .</th></tr>
<tr><td>Resources comprising pages, leaves, sheets, or cards (or images of these)</td><td>Books, printed music, maps, manuscripts, posters, printed serials, sheets of microforms, flashcards, etc.</td><td>Title page, title sheet, or title card (or an image of these)</td><td>A cover/jacket issued with resource, a caption, a masthead, or a colophon (or an image of these) or, if those are unavailable, then another source within resource</td></tr>
<tr><td>Resources comprising moving images</td><td>Film reel, a DVD, a video game, an MPEG video file, etc.</td><td>Title frame(s) or title screen(s)</td><td>If a tangible moving image resource:
(1) a label permanently printed on or affixed to resource,
(2) a container or accompanying material issued with resource,
(3) an internal source forming part of a tangible digital resource (e.g., a DVD menu), or
(4) another source within resource

If an online moving image resource:
(1) textual content,
(2) embedded metadata that contains a title (e.g., metadata embedded in an MPEG video file), or
(3) another source within resource</td></tr>
<tr><td>Other resources</td><td>Digital images, databases, globes, CDs, MP3 files, audio cassettes, objects, puppets, kits, websites, etc.</td><td colspan="2">If a tangible resource:
(1) a textual source on resource itself or a label permanently printed on or affixed to resource (e.g., a label on an audio CD or a model),
(2) an internal source, such as a title screen,
(3) a container or accompanying material issued with resource, or
(4) another source within resource

If an online resource:
(1) textual content,
(2) embedded metadata in textual form that contains a title, or
(3) another source within resource</td></tr>
</table>

- Does this resource contain a new work or a new expression?
- Is it a new manifestation of a previously existing resource?
- Is this resource related to other resources? (e.g., Is it based on another work? Does it continue another work?)
- Is this resource part of a larger resource?

These questions will help organizers to identify important bibliographic relationships and the necessary access points to supplement the basic descriptive data. These topics are addressed in the next chapter.

Common Attributes across Resources Types

Information resources have numerous attributes associated with them. Many attributes are useful in fulfilling the user tasks (i.e., find, identify, select, obtain, and explore). Other attributes, however, may not be impor-tant or necessary. Guided by metadata standards, organizers supply the most essential attributes in resource descriptions. What is considered *most essential*, however, varies among information communities, content standards, and schemas. What may be important to one community may not be important to another. For example, although a publishing company may find the name of the copy editor for a particular project to be important metadata in their recordkeeping system, the same information is not considered significant when describing a book for a library. It all depends on the context. Although metadata may look somewhat different depending on the type of resource being described and the community who creates that description, there are consistencies to be found. In the next sections, attributes common to many different resource types are enumerated and described.

Titles, or, What Is It Called?

Most resources have a title. A *title* is a name that has been assigned to a resource by a creator, contributor, producer, owner, or, perhaps, someone else (e.g., an archivist). Identifying and finding a resource is much easier when the resource is called something. Some titles are descriptive (e.g., *Portrait of Cornelis van der Geest* or *Introduction to Cataloging and Classification*), but some are less so (e.g., *Charity* or *The Naked Gun*), and others may be fairly unhelpful (e.g., *Untitled*). There are many different kinds of titles including, but not limited to, the following:

- **Title Proper**: The main or primary title by which the resource is known (e.g., *The Brief Wondrous Life of Oscar Wao*).
- **Subtitle**: An additional title, subordinate to the title proper, that augments the title proper (e.g., *Tess of the D'Urbervilles: A Pure Woman, Faithfully Presented*).
- **Other Title Information**: An additional title, phrase, or statement that helps to qualify or amplify the title. All subtitles are a form of other title information, but not all examples of other title information are subtitles. For example, some other title information explains the form of the work (e.g., *A Moon for the Misbegotten: A Play in Four Acts*).
- **Devised Title**: The title provided by an archivist for a collection or by another information professional for an untitled resource (e.g., *Paul Hibbet Clyde and Mary Kestler Family Papers*).
- **Parallel Title**: A title proper that is repeated in another language or script (e.g., *Ma Vie en Rose* = *My Life in Pink*).
- **Alternative Title**: A second title for a resource that is joined to the first title with the word *or* (e.g., *Frankenstein, or, The Modern Prometheus*). Both titles together are considered to constitute the title proper of a resource.
- **Collective Title**: An inclusive title that represents a compilation (e.g., *Ten Days That Shook Scotland*) that contains two or more individually titled parts (e.g., 10 essays about key moments in the history of Scottish soccer).
- **Conventional Collective Title**: A collective title referring to the form of the work used for a compilation containing two or more works by a person, family, or corporate body (e.g., *Plays*, *Works*, *Speeches*) or used for two or more parts of a work (e.g., *Selections*).

In addition, a number of other titles exist, such as cover titles, corrected titles, spine titles, and so on. In CCO, rather than creating different elements for different types of titles, an additional element—Title Type—"provides a way to distinguish between the various types of titles (for example, repository title, inscribed title, creator's title, descriptive title)."[14]

When recording titles, a number of issues may arise. Among other questions, organizers might ask: Where are titles found? If there is more than one, which title should be treated as the title proper? How are inaccuracies or typos addressed? Are introductory words or phrases recorded (e.g., Disney presents *Sleeping Beauty*)? If the archivist or cataloger must supply a title, what words, names, or concepts should be included? What if there are changes in a title over time (e.g., changes in a journal's title)? The answers to these questions may vary depending on the content standard used to guide the description.

Editions, or, Which Version Is It?

Once the name of the resource is understood, it is important to determine if the resource has appeared in different versions. An *edition* is a particular version of a resource. Print materials, moving images, sound recordings, online resources, and other resource types might have version information associated with them. An *edition statement* is a declaration of the version; it is usually transcribed exactly as it appears on the resource itself, but it may come from outside sources if needed. In art and cultural heritage collections, *edition* and *state* are elements that are used with art works produced in multiples. *Edition* identifies a specific print in a limited run of identical artworks (e.g., 12/30, indicating this piece is the 12th of 30 identical pieces), and *state* refers to alterations to or variations of the original version of a print or another work of art. Some examples of version metadata include the following:

- Eleventh edition
- 4th ed.
- Household edition
- Version 10.9.3
- Interactive version
- 22/400
- 7th of 10 states

When recording edition or version information, a number of problems or issues may arise. Among other questions, organizers might ask: Where do I find edition information? How do I record statements of responsibility associated with an edition? Should I use abbreviations and numerals? What do these numbers mean? What are the equivalent concepts in other languages? For example, do *edition* and *edición* mean the same thing or are there differences between them?

Dissemination Information, or, Where Did It Come From and When?

Resource descriptions, of all types, typically include information about the publication, distribution, presentation, manufacturing, or production of the resource. In short, our main concerns are:

- Where is the resource from?
- Who or what produced (published, printed, manufactured, etc.) it?
- When was it made available?

The description, therefore, includes information about chronological periods associated with the resource, responsible bodies for the production, and geographic locations from which the resource came.

In library cataloging, for tangible materials such as books and maps, the name of the publisher and the date and location of publication are included, whenever they are available. Additionally, manufacturing, production, and distribution information, as well as copyright dates, may be included if deemed important. For example, in a catalog record for a book, one might see something similar to the following publication statement:

264 _1 Santa Barbara, California : **$b** Libraries Unlimited, **$c** 2017.

In a Dublin Core description for the same resource, one might instead find this:

<dc:publisher>Libraries Unlimited**</dc:publisher>**
<dc:date>2017**</dc:date>**

Aside from the display and encoding differences, the data is abbreviated because Dublin Core does not have a specific element for publication location.

In archives, publisher's names and locations (or those of a manufacturer or a distributor) are not likely to be recorded because archival collections are typically not considered to be published resources (although published materials may be part of the collection). On the other hand, dates are important pieces of data to be included. Dates may be associated with the collection as a whole (e.g., dates of creation, dates of acquisition) or with individual parts of the collection (e.g., dates for a particular series). In addition, locations associated with the collection (where it was created, assembled, accumulated, and/or maintained) and locations associated with those responsible for the collection are included in the description.

There are, as might be expected, numerous issues that can arise when recording dissemination information. Organizers may ask: What if there are multiple places of publication (distribution, manufacture, or production)? How much of the place name should be recorded? What if there are two or more corporate bodies listed as publishers? What is recorded for a self-produced resource? Can an individual person be a publisher? What if no dissemination information is provided at all? What do I do with all of these dates?

In any repository type, date elements, in particular, can be confounding. This is because there are so many different kinds of dates. For example, one might encounter dates associated with

- Publication
- Distribution
- Manufacturing
- Construction
- Copyright
- Production
- Creation
- Design
- Presentation
- Performance
- Modification
- Alteration
- Access
- Activity
- Availability
- Record-keeping activity
- Broadcast
- Display
- Exhibition

In addition to sorting through all the different kinds of dates, recording dates can be challenging. There are exact dates (15 January 1973), single years (1965), open dates (1941-), date ranges (1973-1981), approximate dates (2012?), approximate ranges (early sixteenth century), earliest and latest dates (1500 to 1530), and uncertain or unknown dates with which to contend. When confronted with problems related to dissemination information, one must remember that each metadata standard has its own sets of elements and instructions for their use. How these problems are addressed may vary from standard to standard.

Physical Description, or, What Does It Look Like?

For all tangible resources and for many online resources, organizers are expected to describe the physical or electronic form of the resource. Usually, the organizer creates this metadata after examining the resource (e.g., counting, measuring, skimming). This metadata helps users to understand exactly what is being described. It tells them, for example, if the latest Adele album they are seeking is available as a compact disc, a long-playing record album, a reel-to-reel tape, an mp3 file, or an AAC file. In general, organizers want to know the following information about a resource:

- How many pages, volumes, documents, boxes, discs, objects, etc., are included?

- How big is the resource? What are its measurements?
- What does the resource look like, sound like, feel like, etc.? Is there color or sound, or are there illustrations? Is it analog or digital? What are the techniques used to create it? What is it made of?
- What are the resource's content, carrier, and media type(s)? What is its format?

The physical description of the resource, of course, varies depending on the type of material in hand. For example, a website and an archival collection require different kinds of data. Examples from a library, the web, an archives, and a museum follow:

300 _ _ 1 hand puppet : **$b** red and blue ; **$c** 20 cm + **$e** 1 teacher's booklet (14 pages ; 15 cm)

<**dc:type**>text</**dc:type**>
<**dc:format**>html</**dc:format**>

Extent: 10 boxes
Extent: 45 linear feet, including 200 photographs and 16 maps
Extent: 1 box, 15 folders

Materials and Techniques: pot-metal glass with vitreous paint
Dimensions Description: 63.5 × 71.5 cm (25 × 28 1/8 inches)
  Value: 63.5 **Unit**: cm **Type**: height
  Value: 71.5 **Unit**: cm **Type**: width

Creators, or, Who Is Responsible for It?

Not only do resources need to be described, they also need to be connected with the agents responsible for their creation and development. Organizers, therefore, need to understand who or what is responsible for creating and contributing to a resource. The names and activities associated with creators and contributors are important pieces of information that should always be included in a resource description, if they are knowable and available. In some organizing traditions, this area of the description overlaps with the provision of access points for creators and contributors (discussed in the next chapter).

Creators and contributors may appear as part of the descriptive data and as access points added to the description. This is not, however, duplicate information. There are two different functions being addressed in these two forms of data. When the names of persons, families, and corporate bodies are included in the descriptive data, these are commonly included as part of a statement of responsibility. A *statement of responsibility* names the agents responsible for the intellectual or artistic content of a work, contributions to an expression, performances of the content, revisions of a work, subsequent editions, and so on. A statement of responsibility is transcribed exactly as it appears on the resource. This allows for the resource to be matched precisely to the description, helping to fulfill the user task *identify*. When these agents are *also* recorded as access points, their names are placed under authority control, and are, therefore, standardized. They do not necessarily reflect the form of name appearing on the resource itself, but they are necessary in the catalog for collocation of agents that may use different forms of name at different points in their existences. For example, the title page of a book might provide the information shown in Figure 7.2.

From it, we could derive the following metadata:

Statement of Responsibility: by M. Grieve

Subsequent Statement of Responsibility: with an introduction by the editor Mrs. C.F. Leyle

The authority-controlled access points, however, for the same agents look a bit different:

Creator: Grieve, Maud, Mrs.

Contributor: Leyle, C. F., Mrs., 1890–1957.

A Modern Herbal:

The Medicinal, Culinary, Cosmetic and Economic
Properties, Cultivation and Folk-Lore Of Herbs,
Grasses, Fungi, Shrubs & Trees, With All Their Modern
Scientific Uses With a New Service Index

by M. Grieve

with an introduction by
the editor Mrs. C.F. Leyle

New York
Hafner
1967

Figure 7.2. A Representation of a Title Page.

In each case, the name is somewhat changed in the access point. The creator's name includes a first name and a term of address (Mrs.), and the name is entered in indirect order. In the contributor's name, the order and the spacing are different, and birth and death dates have been added in the access point. Both museums and archives have similar issues in that they may include transcribed data as well as access points under authority control in their resource descriptions.

Other Common Characteristics

In addition to those enumerated above, other characteristics common across resource types and communities include those reflecting the content of works. These may include any of the following: subject headings, thesaurus terms, classification notations, abstracts or summaries, scope and contents notes, tables of contents, genre/form descriptions, and culture, style, and language information. Subject matter and related characteristics are addressed in more detail in later chapters.

Many resources also have relationship and/or contextual information included in the description. This category includes attributes such as series information, accompanying materials, publication history, source and other relations, numbering information, biographical information or corporate history, provenance information, accession information, and custodial history. This data helps users to better understand the connections among resources and relevant background information. As mentioned earlier, the next chapter discusses relationships in more depth.

Finally, every resource may have some administrative or identifying metadata that helps to locate and access the resource. This includes call numbers, shelf addresses, accession numbers, and standard identifiers (such as a Uniform Resource Identifier [URI], Internationalized Resource Identifier [IRI], Uniform Resource Locator [URL], Digital Object Identifier [DOI], International Standard Book Number [ISBN], and so on). The common attributes described in this section are only just scratching the surface of descriptive elements. They are presented, here, at a very high level. More specific information about each of these and other areas can be found in the metadata standards themselves.

CREATION OF METADATA DESCRIPTIONS

Once the nature of the resource has been determined, a description is created by selecting important pieces of data (e.g., title, creator, date) from the information resource, determining certain characteristics about the resource (e.g., size, terms of availability), and then placing those pieces of information in a certain order, usually dictated by a set of rules or conventions for description. These rules or conventions (i.e., *content standards*) are created by different communities, so that those communities can describe appropriately the resources for which they are responsible. Content standards serve as style manuals for metadata, identifying the elements to be included, providing definitions of each element, and sometimes providing rules for exactly what information to include in a description, for the structure of that information, and occasionally for its punctuation and order. Some standards also provide information about where to find the metadata and how to address metadata in different languages or scripts; some address issues related to capitalization, initials, numbers, and dates; and some provide instructions on the use of symbols, diacritical marks, abbreviations, and acronyms, among other topics.

Several of today's metadata schemas are outgrowths of rules that were known as bibliographic schemas or *cataloging codes*. Such rules were essentially content standards, first for records in print retrieval tools and later for records to be entered into online systems. As online retrieval tools came into being, separate standards for the encoding of surrogate records were developed to be used to create online records. The conceptual pieces necessary for online records are, then,

- **Elements**: identification of which types of information (i.e., attributes) are to be included,
- **Content**: metadata values, which may be prescribed with formatting instructions or may be just loosely described in the standard, and
- **Syntax**: encoding for machine manipulation through the use of an encoding standard or a markup language.

A standard may dictate elements only (e.g., Dublin Core), content only (e.g., CCO), or syntax only (e.g., XML schemas). Some metadata standards, however, have been created that combine the conceptual pieces in different ways. The Metadata Object Description Schema (MODS) used by libraries, for example, specifies what elements are available for use (i.e., it is a data structure standard), dictates the content and form of some elements (as does a content standard), and specifies the syntax to use (i.e., it dictates the encoding standard to use). Increasingly, metadata standards address more than one of the various components of metadata outlined in Table 5.1 on page 183.

An information resource's properties may be described using community-specific or schema-associated rules. Several examples of metadata standards from different communities are discussed throughout the rest of this chapter. The emphasis in this chapter is on the content component of each schema discussed. The selection is not exhaustive, but illustrative.

- **Bibliographic and General Metadata Schemas**
    - *RDA: Resource Description & Access*
    - *International Standard Bibliographic Description* (ISBD)

   - *Anglo-American Cataloguing Rules, Second Edition, 2002 revision* (AACR2R), with updates issued in 2003, 2004, and 2005
   - Dublin Core (DC)
   - Metadata Object Description Schema (MODS)
- **Archives Metadata Schemas**
   - *General International Standard Archival Description* (ISAD(G))
   - *Describing Archives: A Content Standard* (DACS)
   - Encoded Archival Description (EAD)
- **Other Domain-Specific Metadata Schemas**
   - TEI (Text Encoding Initiative) Headers
   - *Cataloging Cultural Objects* (CCO)
   - VRA (Visual Resources Association) Core
   - *Categories for the Description of Works of Art* (CDWA)
   - ONIX (Online Information Exchange)
   - Index and bibliography records

Bibliographic and General Metadata Schemas

Metadata standards in the library field were developed long before the term *metadata* was ever uttered. Some of our modern guidelines for description and access can be traced back to rules developed by Anthony Panizzi of the British Library in the first part of the nineteenth century. Cataloging codes, however, existed even before that, with the first national code having been developed in France in the aftermath of the French Revolution. Examples of such cataloging standards discussed here are RDA and AACR2. ISBD, a long-established metadata element set and content standard, is also described, although ISBD's influence has diminished in recent years with the advent of RDA. MAchine-Readable Cataloging (MARC), the standard currently used to disseminate much bibliographic metadata in libraries, is discussed in detail in Chapter 6, as it is an encoding schema. As the need for metadata for other communities, especially for electronic resources in those communities, became apparent in the mid-1990s, the Dublin Core was conceived and developed as a general-purpose schema. MODS, another general-purpose schema, followed in 2002; it was offered by the Library of Congress (LC) as an alternative to Dublin Core and to MARC-encoded AACR2 cataloging records. In Textbox 7.1, another influential cataloging document, IFLA's *Statement of International Cataloguing Principles* (ICP), is described.[15]

IFLA's *Statement of International Cataloguing Principles* (ICP)

Although IFLA's ICP is not a description standard per se, it is an important document in international cataloguing that should be mentioned. The following is a modified excerpt from Daniel N. Joudrey, Arlene G. Taylor, and David P. Miller's 11th edition of *Introduction to Cataloging and Classification*:

> After a series of five meetings held around the world between 2003 and 2007, the IFLA Meeting of Experts on an International Cataloguing Code (IME ICC) disseminated a document called the *Statement of International Cataloguing Principles* (ICP) in 2009, which has since been revised and republished in 2016. The ICP arose from concerns about the future of cataloging that had been fodder for discussions over the previous decade or longer. The purpose of the ICP was to update and expand upon the previous statement of cataloging principles produced by the International Conference on Cataloguing Principles held in Paris in 1961. These "Paris Principles," based largely on the work of Seymour Lubetzky, have long been acknowledged as a guiding force in the development of cataloging codes around the globe concerning the choice and form of access points. From this point into the foreseeable future, the ICP is to be the basis for the international standardization of cataloging, having been built on earlier cataloging traditions and on the foundations of the FRBR and FRAD conceptual models, which were also created by IFLA working groups. The idea is that the Statement of International Cataloguing Principles can be used as a prefatory document that will ensure a degree of consistency among the cataloging codes created by various countries around the world.

The ICP contains 13 basic principles. The following is a brief interpretation of each:

1. **Convenience of the User**: Think of the user first!
2. **Common Usage**: Use the language of everyday users.
3. **Representation**: The description should reflect how a resource describes itself.
4. **Accuracy**: Get the description right!
5. **Sufficiency and Necessity**: Require the data needed to make the resource findable and understandable.
6. **Significance**: Stick with the most relevant attributes to describe the resource.
7. **Economy**: Prefer ways to describe something quickly and efficiently.
8. **Consistency and Standardization**: Standardize as much data as possible, whenever possible.
9. **Integration**: Start with widely applicable rules, rather than specialized rules for specific resource types.
10. **Interoperability**: Work towards the sharing and re-use of data as much as possible.
11. **Openness**: Make data as available as possible. Do not place unnecessary restrictions on it.
12. **Accessibility**: Make it accessible to all. Comply with international standards for accessibility.
13. **Rationality**: Rules should make sense; they should be rational, not arbitrary.

In addition to these general principles, there are sections of the ICP on entities, attributes, and relationships; bibliographic descriptions; access points; objectives and functions of the catalogue; and foundations for search capabilities, among others.

> For more information, please consult: IFLA Cataloguing Section and IFLA Meetings of Experts on an International Cataloguing Code. *Statement of International Cataloguing Principles (ICP)*. 2016 edition by Agnese Galeffi, María Violeta Bertolini, Robert L. Bothmann, Elena Escolano Rodríguez, and Dorothy McGarry. The Hague: IFLA, 2016. http://www.ifla.org/files/assets/cataloguing/icp/icp_2016-en.pdf.

Textbox 7.1. A Brief Overview of IFLA's ICP.

RDA: Resource Description & Access

RDA: Resource Description & Access, published in 2010, is the latest in a long line of descriptive cataloging content standards. Although it has been published in print form, RDA is intended to be accessed primarily through the *RDA Toolkit*, a subscription-based electronic tool. The toolkit includes the text of the content standard, along with other features such as workflows, mappings, additional examples, archives of superseded instructions, and links to other related resources. In this section, some features of the RDA Toolkit are described as part of the discussion of the features of RDA itself.

RDA is the intended replacement for the second edition of the *Anglo-American Cataloguing Rules* (AACR2). After a period of "thorough and rigorous" testing, the U.S. national libraries—LC, the National Library of Medicine, and the National Agricultural Library—implemented it in early 2013.[16] Numerous academic, special, and public libraries in the United States also have implemented RDA. National (or other notable) libraries in Australia, Austria, Canada, Catalonia, Czech Republic, England, Finland, Germany, Iceland, Ireland, Italy, Latvia, Malaysia, the Netherlands, New Zealand, the Philippines, Singapore, Slovakia, and Switzerland have adopted RDA, and libraries in Denmark, Estonia, Lithuania, Norway, and Spain are looking at possible implementations as well.[17] At this time, six translations of the content standard are included in the *RDA Toolkit*: Catalan, Finnish, French, German, Italian, and Spanish. A print version of a Chinese translation is available, and other translations, such as the one in Norwegian, are under way or will be developed over time.

At the beginning of the twenty-first century, many catalogers felt that a new approach to describing resources was needed. AACR2 was showing its age; its handling of online and other electronic resources was awkward at best. Cataloging experts felt there was need for a principles-based approach to cataloging (rather than establishing rules for every possible situation—Andrew Osborn's legalistic approach to cataloging[18] discussed in Chapter 2)—but that it must be established without completely dismantling the work that had come before it.

RDA is based on earlier cataloging standards. It attempts to create a global approach to describing information resources, built upon Anglo-American cataloging traditions, two of IFLA's conceptual, functional-requirements models (FRBR and *Functional Requirements for Authority Data* [FRAD][19]), the harmonization of those models (LRM[20]; discussed below), and IFLA's statement of principles (ICP[21]). From FRBR and FRAD, RDA employs the underlying entity-relationship model, the FRBR user tasks, many of the entities identified in FRBR, and some of the numerous attributes and relationships detailed in the models. From the ICP, RDA embraces the basic principles enumerated in the statement, such as the principles of *representation* and *convenience of the user*. RDA combines the conceptual models and the principles into a foundation on which to build a standard that allows for the greater use of cataloger's judgment and the creation of data that is more compatible with future technologies and machine processing. RDA is also based on the previous content standard, AACR2, with which it is intended to be backward compatible; many of the actual cataloging instructions in RDA read the same as they were found in AACR2, but the order of those instructions has changed dramatically.

RDA is intended to be more international in scope and *schema-neutral*. In other words, it is not meant be as Anglo-centric as the previous content standard, and it is meant to be implemented independently from any particular encoding standard or display format. Rather than being tied to the English language, individual libraries preparing bibliographic descriptions can make choices about the scripts and languages used in supplied data and access points, the transliteration approaches used, and the calendars and numeric systems employed. Being schema-neutral, RDA is not specifically tied to MARC, ISBD, an XML schema, a linked data format such as BIBFRAME, or any other system. Thus, the instructions in RDA can be used to formulate descriptive metadata that may be communicated using any format available. For example, RDA-compliant metadata easily could be placed in a MARC record or it could be placed in an XML-encoded Dublin Core description. RDA also does not concern itself with how the data should be displayed. In the United States and some other Anglo-American countries, we often use the punctuation from ISBD to create groupings of different elements in our cataloging data. RDA does not require this type of display, but many libraries have chosen to implement it in this manner based on LC's decision to retain ISBD punctuation in its MARC records. It was announced in 2016, however, that the Program for Cooperative Cataloging (PCC) was recommending strategies for the elimination of ISBD punctuation from MARC-encoded bibliographic metadata.[22] For more on ISBD, see below.

As a content standard, RDA provides guidelines on how to describe a resource. In previous content standards, the focus has been primarily on creating descriptive data for a resource and on the choice and form of access points. In RDA, the description of other types of entities is included. There are individual chapters, for example, that address how to describe persons, families, and corporate bodies. These chapters list myriad attributes that can be used to create authority data for each of these types of agents. It also provides information about identifying relationships, often, but not always, through the creation of access points.

The FRBR model's entities and relationships act as the skeleton for the content standard. At the time of this writing, however, it has been announced that RDA's overall structure will change based on revisions coming from the upcoming IFLA LRM.[23] Preliminary discussions indicate that there will be individual chapters for most of the individual LRM entities (agent, collective agent, expression, item, manifestation, nomen, person, place, time-span, and work); the exception is *res*, which is not expected to have its own chapter and is likely to be renamed *RDA entity* in the content standard. In addition, there will be one or more chapters containing general instructions on describing entities and recording data; a chapter on aggregates; a chapter on pre-cataloging decisions; and chapters on families and corporate bodies (entities in RDA that are not included in LRM). The new structure is scheduled to be implemented in 2018. Please consult RDA directly to see changes that have occurred.

In the pre-LRM iteration of RDA, the content standard begins with a general introduction, with the rest of the chapters divided up into 10 sections:

- Section 1: Recording Attributes of Manifestation & Item
- Section 2: Recording Attributes of Work & Expression
- Section 3: Recording Attributes of Person, Family, & Corporate Body
- Section 4: Recording Attributes of Concept, Object, Event, & Place
- Section 5: Recording Primary Relationships between Work, Expression, Manifestation, & Item
- Section 6: Recording Relationships to Persons, Families, & Corporate Bodies
- Section 7: Recording Relationships to Concepts, Objects, Events, & Places
- Section 8: Recording Relationships between Works, Expressions, Manifestations, & Items
- Section 9: Recording Relationships between Persons, Families, & Corporate Bodies
- Section 10: Recording Relationships between Concepts, Objects, Events, & Places

As can be seen, Sections 1–4 are dedicated to describing entities (i.e., works, expressions, manifestations, items, different agents, and subjects), and Sections 5–10 are focused on recording relationships. Most sections contain between one and five chapters. For example, at the time of this writing Section 1 contains the following chapters:

- Section 1: Recording Attributes of Manifestation & Item
    - 1: General Guidelines on Recording Attributes of Manifestations and Items
    - 2: Identifying Manifestations and Items
    - 3: Describing Carriers
    - 4: Providing Acquisition and Access Information

The sections related to subjects, however, are mostly empty. As stated, changes in the structure of RDA are expected, but the *exact* changes are unknown at the time of this writing.

The chapters are broken down into logical components based on the nature of the chapter. Chapter 2: Identifying Manifestations and Items, for example, begins with three introductory sections focused on the chapter's scope, the basis for identifying the resource, and sources of information. What follows is an enumeration of the major attributes used to describe manifestations and items:

- Title
- Statement of Responsibility
- Edition Statement
- Numbering of Serials
- Production Statement
- Publication Statement
- Distribution Statement
- Manufacture Statement
- Copyright Date
- Series Statement
- Mode of Issuance
- Frequency
- Identifier for the Manifestation
- Preferred Citation
- Note on Manifestation

- Custodial History of Item
- Immediate Source of Acquisition of Item
- Identifier for the Item
- Note on Item

Each attribute may contain a single metadata element or two or more sub-elements, each with one or many instructions. For example, under Publication Statement there are entries for

- Basic Instructions on Recording Publication Statements
- Place of Publication
- Parallel Place of Publication
- Publisher's Name
- Parallel Publisher's Name
- Date of Publication

For each element, RDA provides a definition and the sources of information to consult, followed by basic instructions for recording the element.

Some elements are identified in RDA as core. A *core element* is one that is required in all descriptions (if the information is applicable and discernible). Some elements are always core, but others are considered *core-if*, meaning that the element is required only in certain circumstances. For example, the element *extent* is core, but only if the resource is complete or if the total extent is known. In the RDA Toolkit, a label identifies individual core elements in the instructions.

Carrier Type

CORE ELEMENT

At the time of this writing, RDA identifies 19 manifestation-level elements as core. These are

- Title proper
- Statement of responsibility relating to title proper
- Designation of edition
- Designation of a named revision of an edition
- Numeric/alphabetic designation of first issue or part of sequence
- Chronological designation of first issue
- Numeric/alphabetic designation of last issue
- Chronological designation of last issue
- Date of production
- Place of publication
- Publisher's name
- Date of publication
- Title proper of series
- Numbering within series
- Title proper of subseries
- Numbering within subseries
- Identifier for the manifestation
- Carrier type
- Extent

There are, of course, additional core elements for works, expressions, persons, and the other entities described using the instructions in RDA. The complete list of RDA core elements can be found in the instructions at RDA 0.6. It should be remembered that *core elements* represent the bare minimum for description. Conscientious catalogers record *all* the data needed to make a resource findable, identifiable, understandable, and so on.

An *instruction*, unsurprisingly, is a statement that tells the cataloger what to do when transcribing or otherwise formulating the metadata. An example of a multipart instruction appears in Figure 7.3. Illustrative examples typically follow instructions or parts of instructions. In the Toolkit, examples are placed in textboxes to ensure that they are noticeable on the screen.

You may notice that an exception appears in Figure 7.3. *Exceptions* describe situations where the cataloger is expected to deviate from standard practice; exceptions are usually made for specific situations or for particular resource types. Exceptions are always applied whenever the circumstances warrant it. In addition to exceptions, catalogers might also encounter alternatives and options. *Alternatives* provide a different approach to what was specified in the immediately preceding instruction. Options come in two flavors: optional additions and optional omissions. Obviously, *optional additions* allow catalogers to supplement the information required by the previous instruction with more data; *optional omissions* allow catalogers to omit some information called for in the preceding instruction when it is not absolutely essential to identify the resource. Whether to apply alternatives and options is based on cataloger's judgment, local policies, or an existing policy statement (see below). In the Toolkit, exceptions, options, and alternatives are labeled with italicized lettering and a vertical bar in the left margin, as can be seen in Figure 7.3.

Also seen in Figure 7.3, next to the instruction number and its caption, is an icon that reads *LC-PCC PS*. In the Toolkit, this is a link to a policy statement. A *policy statement* contains an interpretation of an instruction, guidance on applying alternatives and options, or a supplemental policy that applies to an instruction. Policy statements also identify additional core elements. The PS icon in the figure is a link to the *Library of Congress-Program for Cooperative Cataloging Policy Statements* (LC-PCC PSs). Many U.S. libraries that use RDA also follow the LC-PCC PSs so that their cataloging is consistent with that of LC and PCC libraries. This means that catalogers at these libraries not only include RDA's core elements in their descriptions but also the additional elements that LC and the PCC have determined should be required (if applicable and discernible). For example, the LC-PCC PSs add 19 more core elements for the description of manifestations:

2.12.2.3 Recording Title Proper of Series LC-PCC PS

Record the title proper of series by applying the basic instructions on recording titles at 2.3.1.

EXAMPLE

Bartholomew world travel series

Great sacred choruses

Allstate simulation film library

. . .

Record an alternative title proper of series as part of the title proper of series.

If the title proper of series includes numbering as an integral part of the title, transcribe the numbering as part of the title proper of series.

EXAMPLE

Publication #122 of the Social Science Education Consortium

The twenty-sixth L. Ray Buckendale lecture

Cuaderno número G del instituto

Exception

If:

the resource being described consists of two or more issues or parts

and

numbering that is an integral part of the title proper of series differs from issue to issue or part to part

then:

omit the numbering from the title proper of series. Use a mark of omission (...) to indicate such an omission. Record the numbering as numbering within the series (see 2.12.9).

EXAMPLE

Publication ... of the Indiana University Research Center in Anthropology, Folklore, and Linguistics

Figure 7.3. An Example of an RDA Instruction. (Based on the layout of RDA 2.12.2.3 in the RDA Toolkit)

- Parallel title proper
- Other title information

- Earlier title proper
- Later title proper
- Key title
- Abbreviated title
- ISSN of series
- ISSN of subseries
- Mode of issuance
- Frequency
- Note on title
- Note on issue/part
- Media type
- Dimensions
- Layout
- Digital file characteristic
- Note on changes in carrier
- Restrictions on use
- Uniform Resource Locator

In addition to the links to the LC-PCC PSs, there are also icons in the Toolkit that link to policy statements produced by the national libraries of Australia, Canada, Finland, Germany, Sweden, and the United Kingdom, as well as those provided by the Music Library Association.

In order to start a basic bibliographic description, catalogers transcribe or record values in each applicable RDA element. This might be done using a template found in the cataloging module of an integrated library system, in a template used in a bibliographic utility such as OCLC Online Computer Library Center, or by typing out the metadata on a card. In RDA, four different approaches to recording data are possible; this is referred to as the *four-fold path*. Some of these paths, however, will not apply to every data element in RDA (e.g., you would not likely use an access point or an IRI for a copyright date). The four methods include:

- **Unstructured descriptions**: An unstructured description is an informal string of narrative data. It could be a free-text note or a straightforward transcription of data directly from the source. "A significant expansion of the tenth edition published in 2006 by Arlene G. Taylor" is an unstructured description, as is "edited by Susan Hill" in a statement of responsibility. Unstructured descriptions may be full or partial.
- **Structured descriptions**: A structured description is a formal string of organized or controlled data. Sub-types include the following:
  - **Authorized Access Points**: An AAP is a human-readable, established string of alphanumeric characters used to represent an entity, usually made up of the entity's preferred name, title, or a combination of both. AAPs are discussed extensively in Chapter 8. They may be used for agents, works, or expressions (manifestations and items typically do not have established AAPs). For example, an AAP for a work is **Garner, Bryan A. Modern American usage**, a name/title access point.
  - **Aggregated descriptions**: An aggregated description is an amalgamation of formal data elements used to create a description of the resource. For example, an aggregated description may be brought together using ISBD punctuation (see below) in order to provide a brief accounting of a resource. It may be a full or partial description, but it should provide enough information to definitively identify the resource:

    Introduction to cataloging and classification / Daniel N. Joudrey, Arlene G. Taylor, and David P. Miller. – Eleventh edition. – Santa Barbara : Libraries Unlimited, 2015. – ISBN 978-1-59884-856-4

- **Identifiers**: An identifier is a human-readable character string uniquely associated with an entity, or with surrogates for entities (such as authority records). For example, **n 79151500** is a standard identifier attached to an authority record for a person; it can be used to represent that person. For human understanding, however, the AAP for that person, **Wharton, Edith, 1862–1937**, could be included as well.
- **Uniform or Internationalized Resource Identifiers (URIs/IRIs)**: In a linked data implementation, a URI or IRI can be used to indicate relationships among entities (e.g., the compact URI *rdam:P30001* means *has carrier type*).

For simplicity's sake, catalogers may simply work their way through the elements until all the important attributes are addressed. The order of the elements in RDA is similar to that found in ISBD and AACR2 (see below). RDA chapters 2–4 cover most of the basic descriptive elements at the manifestation and item levels, and Chapter 7 addresses numerous content-related elements (at the work and expression levels), all of which form the body of the bibliographic description. Content elements include, but are not limited to, characteristics such as

- Nature of the Content
- Coverage of the Content
- Intended Audience
- Dissertation or Thesis Information
- Summarization of the Content
- Place and Date of Capture
- Language of the Content
- Accessibility Content
- Illustrative Content
- Supplementary Content
- Colour Content
- Sound Content
- Aspect Ratio
- Duration
- Award

Later chapters of RDA focus primarily on authority data and access points, which are needed for complete catalog records; these elements are discussed in Chapter 8 of this text.

RDA is not used in isolation. As discussed above, policy statements might be followed, although this is not required. Following LC practices, libraries in the United States implementing RDA are doing so in coordination with MARC as the encoding standard (for now) and with the system of punctuation established in the ISBD as a display standard (for now). This should make it very clear that library cataloging is a complex process that requires several different standards in order to be complete (and this does not even begin to address the subject analysis aspects of cataloging!). For an example of RDA data used to describe a resource, please see the template provided in Figure 7.4. For an example of RDA cataloging implemented using MARC and ISBD, please see Figures 6.3 and 6.4 on pages 258–59. For a comprehensive overview of the process of descriptive cataloging using RDA, please see Joudrey, Taylor, and Miller's *Introduction to Cataloging and Classification*.[23]

RDA	Element	Metadata Content	MARC	ISBD
Manifestations and Items				
2.3.2	Title proper (T)	**Introduction to cataloging and classification**	245 $a	n/a
2.4.2	Statement of responsibility for title proper (T)	**Daniel N. Joudrey, Arlene G. Taylor, and David P. Miller**	245 $c	/
2.5	Designation of edition (T)	**Eleventh edition**	250 $a	n/a
2.8.2	Place of publication (T)	**Santa Barbara, California**	264 #1 $a	n/a
2.8.4	Publisher's name (T)	**Libraries Unlimited**	264 #1 $b	:
2.8.6	Date of publication	**[2015]**	264 #1 $c	,
2.12.2	Title proper of series (T)	**Library and information science text series**	490 $a	n/a
2.13	Mode of issuance	**Single unit**	Ldr/07 Value=m	
2.15	Identifier for manifestation	**9781598848571 (hardback : acid-free paper)** **9781598848564 (pbk : acid-free paper)**	020 $a	n/a
Carriers				
3.2	Media type	**Unmediated**	337	n/a
3.3	Carrier type	**Volume**	338	n/a
3.4	Extent	**xxv, 1048 pages**	330 $a	n/a
3.5	Dimensions	**27 cm**	300 $c	;
Works and Expressions				
6.9	Content type	**Text**	336	n/a
Content				
7.10	Summarization of the content	**This new edition reintroduces the topic of library cataloging from a modern perspective. This textbook delineates the new cataloging landscape; shares a principles-based perspective; provides introductory text for beginners and intermediate students; emphasizes descriptive and subject cataloging, as well as format-neutral cataloging, and covers new cataloging rules and RDA**	520 $a	n/a

7.12	Language of the content	**English**	008/35-37 546	n/a
7.15	Illustrative content	**Illustrations**	300 $b	:
7.16	Supplementary content	**Includes bibliographical references (pages 1001-1022) and index.**	500 $a 504 $a	n/a
Persons, Families, and Corporate Bodies (PFCs) Associated with Resource				
19.2	Principal or First-named Creator	**Joudrey, Daniel N., author.**	1XX $a	n/a
19.2	Additional creators	**Taylor, Arlene G., 1941- author.** **Miller, David P. (David Peter), 1955- author**	7XX $a	n/a
Related Resources				
25.1	Related work	**Cataloging in context -- Development of catalogs and cataloging codes -- Underlying principles and conceptual models -- Resource Description and Access (RDA) basics -- Manifestations and items ... [*Table of contents continues*] ... Appendix B: ICC11 RDA book template.** **Based on (work): Taylor, Arlene G., 1941- Introduction to cataloging and classification.** **Library and information science text series.**	500/505/7XX/490	
26.1	Related expression	**Significant expansion of: Introduction to cataloging and classification. Tenth edition / Arlene G. Taylor ; with the assistance of David P. Miller.**	500/7XX	
Other Needed Metadata				
Subject headings and classification data would also be included in the final metadata record.				

Key: **RDA**: RDA instruction number
Element: RDA element name; includes *Core*, *Core-if*, *LC Core*, and *ICC11 Core* for books
(T): Transcribed data
Metadata content: Place to record the data
MARC: Typical MARC fields and subfields
ISBD: Preceding ISBD punctuation when used in MARC

Figure 7.4. An Example of RDA Data Recorded in a Template.

International Standard Bibliographic Description (ISBD)

Around the world, ISBD has long been known as an internationally agreed upon cataloging standard that has been used as a format for the creation and sharing of bibliographic data.[25] ISBD was designed in the early 1970s to facilitate the international exchange of cataloging records by (1) standardizing the elements to be used in the description, (2) assigning an order to these elements, and (3) specifying a system of symbols to be used in punctuating the elements. Actually, there have been several ISBDs based upon different resource types.[26] Previously there were separate ISBDs for monographs, rare (antiquarian) materials, serials, continuing resources, cartographic materials, electronic resources, nonbook materials, and printed music. Efforts to consolidate these separate ISBD standards began in 2003, and a loose-leaf edition called the "Preliminary Consolidated Edition" was released in 2007. The consolidated edition was finalized and published in 2011. Translation efforts and harmonization with RDA are ongoing initiatives of the IFLA ISBD Review Group.[27]

When ISBD was adopted as an international standard, it was expected that national cataloging agencies would incorporate it into their national cataloging rules. It has been widely adopted for use in many countries and incorporated into several sets of national cataloging rules, including AACR2. Joudrey, Taylor, and Miller state

> In countries without a national content standard, ISBD provides order out of what could be chaos. In countries where RDA has been implemented, however, the role and importance of ISBD has been somewhat diminished because RDA has not adopted most of the substantial features of ISBD. Largely, ISBD has been relegated in RDA to a system of punctuation with which online catalogs display their data.[28]

As more libraries in more countries around the world adopt RDA, it is likely that ISBD's significance will continue to diminish, and perhaps one day it will disappear altogether from modern cataloging. As mentioned above, the PCC is discussing strategies to eliminate ISBD from MARC-based cataloging.[29] ISBD's historical significance, however, should not be forgotten.

When used as a descriptive cataloging standard (e.g., for libraries still using catalog cards), ISBD requires that an information resource be totally identified by the description, independent of any access points. It contains nine areas of description:

Area 0—Content form and media type
Area 1—Title and statement of responsibility
Area 2—Edition
Area 3—Material or type of resource specific details
Area 4—Publication, production, distribution, etc.
Area 5—Physical description
Area 6—Series
Area 7—Notes
Area 8—Resource identifier and terms of availability

There are two things to consider when using ISBD for creating resource descriptions. First, ISBD punctuation is prescribed, and it precedes and predicts the data element that comes next. For example, a space-slash-space in Area 1 says that the statement of responsibility is coming next. Second, each area contains more than one element, so the order of data is prescribed. For example, in Area 1 the prescription for content and punctuation is

Title : other title information / 1st statement of responsibility ; subsequent statement of responsibility.

In the above example, it should be noted that the prescribed punctuation is both preceded and followed by a space.

Area 0 identifies the form of the content and type of carrier that contains the resource being described. It comprises three elements: the *content form*, the *content qualification*, and the *media type*. The content form is a "fundamental form or forms in which the content of a resource is expressed."[30] It is mandatory and is populated with a term(s) from a prescribed list:

- Dataset
- Image
- Movement
- Music
- Object
- Program
- Sounds
- Spoken Word
- Text
- Multiple Content Forms

- Other Content Form

The content qualification acts as an expansion of the content form, if it is appropriate for the resource being described. It specifies "the type, presence or absence of motion, dimensionality, or the sensory nature" of the resource.[31] Its value, too, comes from a controlled list of terms found in ISBD (e.g., *aural*, *gustatory*, *olfactory*, *tactile*, and *visual* are listed under sensory qualifiers). The media type provides a specification of the carrier, which reflects the format of the storage medium or housing, as well as any device necessary to interact with the resource. ISBD provides 10 media types:

- Audio
- Electronic
- Microform
- Microscopic
- Projected
- Stereographic
- Unmediated
- Video
- Multiple Media
- Other Media

The following is an example of metadata recorded in Area 0:

Music (performed) : audio

Area 1 contains the *title proper* assigned to the information resource by persons responsible for its existence. There may be more than one type of title, and there may be other information necessary to the understanding of the title (e.g., information about the place and date of a conference). The *statement of responsibility* element of Area 1 contains the names of agents responsible for the intellectual content of the resource, but not necessarily those responsible for presentation and packaging. For example, the name of the artist performing on a music CD would be included here, but not the name of the company that manufactured the CD. An example of Area 1 for a sound recording might look like this:

Army of me : remixes and covers / Björk.

Area 2 contains a statement about the *edition* or *version* of the resource represented in the description being created. It might be a new edition of a work, a new version of a software package, or a version of a work that is put out for a particular geographic area (e.g., the city edition of a news-paper that serves a region). Area 2 also may contain a statement of responsibility, this one relating only to the resource being described (e.g., a person who has worked on the edition in hand but did not work on earlier editions).

Sixth edition / revised by Richard L. DeGowin.

Area 3 contains data that are unique to particular classes of resource. ISBD currently gives rules and examples only for serials, mathematical data for cartographic resources, and music format statements for notated music. For example, in describing serials, it is important to identify the date and the volume number of the first issue of the serial.

Vol. 1, no. 1 (Fall 1980)–

Area 4 contains dissemination information, including the name of the publisher, producer, and/or distributor that is responsible for the issuing and release activities associated with the resource, along with the geographic location of the publisher, producer, or distributor. An important, broadly applicable element in this area is the date of public appearance of the resource. Area 4 is also used for data having to do with physical manufacture of a resource. Such data is given separately from the production data if manufacturing is the work of a separate entity.

New Haven : Yale University Press, 2008.

Area 5 contains a physical description of an information resource. The physical description includes the extent of the resource given in terms of what kind of item it is (e.g., 2 sound discs, 365 p., 4 videocassettes), the dimensions (often height, but also sometimes width, diameter, etc.), and other physical details such as information about illustrations or about material from which an object is made. Area 5 is also used to describe *accompanying material,* which is a physically separable part of the resource that is issued (or intended to be issued) at the same time.

xxxiii, 535 p. : ill., maps ; 25 cm + 1 answer book.

Please note that the example above contains abbreviations. Using abbreviations for *pages* and *illustrations* is recommended practice in ISBD; in RDA cataloging, however, these abbreviations are not allowed.

Area 6 contains the title of a larger bibliographic resource of which the resource is part. This might be a series, subseries, or multipart monographic work. A *series* can be a group of separate works that are related in subject or form and/or are published by the same entity (e.g., *Library and Information Science Text Series*). There may also be information other than the title: a series title can have the same kinds of additional title information as does the title proper in Area 1; and a series may have statements of responsibility that relate only to the series. If a series has an ISSN (International Standard Serial Number), it may appear in Area 6. Numbering of the resource within the larger resource set is also given in Area 6.

(Dictionary of literary biography, 1096-8547 ; v. 379).

Area 7 contains notes relating to the resource being described. Notes may, to name a few, describe the nature, scope, or artistic form of the work; give the language of the text; identify the source of the title if there is no chief source of information; or explain relationships of this work to others. This is the most free-form of the areas.

Includes index.

Originally published by Henmar Press as individual pieces; original copyright dates: 1960, 1977.

Illustrations colored by hand.

Area 8 contains a *standard number* that is used as a resource identifier. Typically, this is defined as a designation assigned by a publisher, or a number recognized as an international standard—which at the moment is only the International Standard Book Number (ISBN) or the ISSN. The area may be repeated if more than one identifier is considered to be important to users. The area also contains information about the terms of availability of the resource (e.g., it is free to members but others must pay, or it is unavailable to the public for a certain number of years).

9780205430826 (pbk.)

Finally, we should say a few words about the formats that ISBD records take. In the ISBD standard, each area is to be set off from the next area by a point-space-dash-space or by the starting of a new line or paragraph. British practice in creating printed catalogs has been to have the areas follow one after another; but American practice has been to create cards and other printed catalogs by having Areas 0, 1, 5, 7, and 8 each begin a new paragraph, and if there is more than one note in Area 7 each note begins a new paragraph (see Figure 3.4 on page 109). In most instances, unless a card catalog is still in use, format is a moot point because the data created using ISBD is placed into MARC records. Displays that are based on MARC records seldom use the catalog card format; they are much more likely to have labels (e.g., **TITLE:** for Area 1). For more information on ISBD, please see the ISBD documentation found on IFLA's website[32] and Joudrey, Taylor, and Miller's *Introduction to Cataloging and Classification*.[33]

Anglo-American Cataloguing Rules, Second Edition (AACR2)

Until 2010, the *Anglo-American Cataloguing Rules, Second Edition*[34] was *the* descriptive cataloging standard for more than 30 years in the United States and other Anglo-American countries. As discussed in Chapter 2, it has had numerous revisions since it was first published in 1978. Major revisions appeared in 1988, 1998, and 2002, with updates continuing through 2005, when the development of a new content standard—originally called *AACR3*, but changed to *RDA*—took precedence. If RDA has replaced AACR2, why is the earlier content standard still being addressed in this book? At this time, not all libraries have moved over to RDA. For a variety of reasons, such as financial concerns, philosophical differences, and/or simple intransigence, some libraries have chosen to continue using AACR2 rather than make the switch. Brief coverage of AACR2 may help newer organizers understand some of the legacy data they may encounter.

For decades, implementation of AACR2 in the United States was dominated by LC, and anyone using this set of rules also consulted the decisions about how LC catalogers interpreted and applied AACR2, as found in the *Library of Congress Rule Interpretations* (LCRI).[35] LC is no longer updating the LCRIs; their content standard is now RDA, and they have replaced the LCRIs with the previously discussed LC-PCC PSs that interpret and supplement RDA instructions. The last version of the LCRIs, published in 2010, is still available through LC's subscription cataloging documentation service *Cataloger's Desktop*.

Much of the impetus for creating a new content standard stemmed from dissatisfaction with AACR2. AACR2 has been criticized for inadequate rules for more complex resources that consist of multiple types of material, including online resources. Other differences exist, especially in terms of access points, but the major differences between RDA and AACR2 descriptions are listed in Table 7.2. For more information on AACR2, please see the 10th edition of Taylor's *Introduction to Cataloging and Classification*,[36] as well as the third edition of *The Organization of Information* by Taylor and Joudrey.[37]

The Dublin Core (DC)

The Dublin Core[38] (shortened form of the Dublin Core Metadata Element Set [DCMES]) was born out of a workshop on metadata semantics held in Dublin, Ohio, in 1995. "At this event, called simply the 'OCLC/NCSA Metadata Workshop,' more than 50 people discussed how a core set of semantics for Web-based resources would be extremely useful for categorizing the Web for easier search and retrieval. They dubbed the result 'Dublin Core metadata' based on the location of the workshop."[39] The idea was to create an internationally agreed-upon set of metadata elements that could be completed by the creators of electronic documents, particularly web resources, which seemed particularly troublesome at the time.[40] The participants in the annual workshops and conferences that have developed and expanded DC are experts from many different fields (e.g., publishers, computer specialists, librarians, information scientists, software producers, text-markup experts). Therefore, it is a cross-domain standard and can be the basis for metadata for any type of resource in any field. It is meant for simple description at low cost; it provides just enough metadata to discover a resource, rather than an extensive or perfect description. It has been approved as ANSI/NISO Standard Z39.85–2012,[41] and ISO standard 15836.[42] It is used internationally, and a number of application profiles have been developed for specific domain applications that use the basic DC as a starting point. It can also be used as an exchange format between institutions that use differing metadata approaches.

Table 7.2. Key Differences between RDA and AACR2.

AACR2	RDA
Descriptions contain a *general material designation* (GMD) to provide an early warning to the user that a resource is in a nonbook format. The GMD is provided immediately after the title proper in square brackets.	Descriptions contain three elements that replace the GMD: • *Content Type* • *Media Type* • *Carrier Type*
Often, the description requires the use of abbreviations (e.g., ill., ed., col., port., min., in., ft., b., d.), including Latin abbreviations, such as *et al.*, *s.l.*, and *s.n.* (for *et alia*, *sine loco*, and *sine nomine*).	In descriptions, RDA restricts the use of all abbreviations. Only abbreviations that appear on the resource may be transcribed. Typically, only abbreviations of dimensions and duration are recorded.
AACR2 employs the *rule of three*, meaning that if a resource has four or more names listed for any one activity in a statement of responsibility (or as access points), only the first name gets included in the description. For example, "by Lois Lane . . . [et al.]."	RDA allows the description to contain as many names in the statement of responsibility as deemed appropriate. They provide an option to limit the names and to be selective of which to include. For example, "by Lois Lane, Clark Kent, Jimmy Olsen, Lana Lang, Perry White [and twelve others]."
AACR2 often encourages manipulating basic descriptive data. For example, typos are corrected in titles, words are omitted from statements of responsibility, and so on.	RDA's principle of representation means catalogers transcribe exactly what is seen on the resource without altering it.
AACR2 is used to describe the item in hand. Properties of the intellectual or artistic work are also addressed.	RDA is based on the FRBR model and addresses not only works and items but is also used to describe manifestations, expressions, persons, families, corporate bodies, and places.

AACR2 uses the following terms:	RDA replaces those terms with:
• Heading	• Authorized access point (AAP)
• Author	• Creator
• Physical description	• Carrier description
• Chief source of information	• Preferred sources of information
• Main entry	• The authorized access point for the creator of a work (the primary or first-named creator)

The Dublin Core Metadata Initiative (DCMI), which is now a project of the Association for Information Science and Technology (ASIS&T),[43] oversees the development of the standard. In its early years DC had been implemented primarily using HTML, but more recently it has been implemented using RDF and XML as well.[44] Templates have been developed that anyone can use to create DC metadata. Such templates can be filled in and previewed, and then HTML-formatted or XML-formatted data can be returned to the user's screen to be copied and pasted into a document header.

The DCMES consists of 15 elements:

- **Contributor**: agent(s) responsible for significant contributions to the content of the resource (e.g., editors, illustrators); the contributions, however, are secondary to those named in the *Creator* element
- **Coverage**: identification of spatial locations (i.e., physical regions), temporal periods (i.e., time periods reflected in the subject matter), or jurisdictions (e.g., named administrative entities) relevant to the content of the resource
- **Creator**: agent(s) *primarily* responsible for creating the content of the resource (e.g., authors, researchers, artists, composers)
- **Date**: date(s) of event(s) in the lifecycle of the resource (such as creation date, availability date, or date of revision); it is recommended that a formal encoding scheme, such as the W3CDTF profile of ISO 8601 (Date and Time Formats), be used[45]
- **Description**: textual statement(s) of the content of the resource; this could be an abstract, a table of contents, a free-text account, or a combination of these
- **Format**: designation(s) of the physical medium, file format, or dimensions of the resource, such as the size or duration; recommended best practice is to select a value from a controlled vocabulary such as the Internet Media Types (also known as *MIME types*)[46]
- **Identifier**: a string or number that uniquely identifies the resource (e.g., URI, URL, DOI, or ISBN); recommended best practice is to conform to a formal identification system
- **Language**: indication of the language(s) of the content of the resource; recommended best practice is to use the language tags defined in RFC 4646[47]
- **Publisher**: name(s) of the agent(s), institution(s), or repository responsible for making the resource available (e.g., person, publishing house, university)
- **Relation**: reference(s) to related resource(s) along with an indication of their relationships to the described resource (e.g., a work that the described resource is a version of, is part of, is referenced by); recommended best practice is to use a string or number from a formal identification system to identify the referenced resource (e.g., URI, IRI, URL, DOI, ISBN)
- **Rights**: statement, link, or identifier that gives information about rights held in and over the resource (e.g., statement about intellectual property rights)
- **Source**: the related resource from which the present one is *derived* in whole or in part; as with *Relation*, recommended best practice is to use a string or number from a formal identification system to identify the referenced resource (e.g., URI, IRI, URL, DOI, ISBN)
- **Subject:** topic(s) reflected in the resource; use of controlled vocabularies and formal classification schemes is encouraged
- **Title**: the name(s) of the information resource
- **Type**: designation(s) of the nature or genre of the content of the resource; recommended best practice is to select a value from a controlled vocabulary such as the DCMI Type Vocabulary (e.g., Dataset, Event, MovingImage, Text)[48]

The general principles for using the DCMES include the following:

- the core set can be extended with further elements needed by a particular community,
- all elements are optional,
- all elements are repeatable,
- any element may be refined by qualifiers, but
- qualifiers cannot change the meaning of the element.

Additionally, DC embraces what is referred to as the *one-to-one principle*. This simply means that a description should refer to only one resource. Each element should be focused on the resource being described, not entities associated with that resource (e.g., the creator). The form of the content of each element, however, is not prescribed by Dublin Core. This means that those creating the descriptions can, in fact, do what they like. Although best-practice recommendations may be made, there is no agency enforcing compliance with these principles or other recommendations.

For a time there were two camps among the DCMI community, each with strong advocates. The Minimalist camp wanted just the 15 elements with no qualifiers. The Qualifiers camp insisted that sub-elements or *refinements* are both useful and necessary. In a sense, both have won. The NISO standard contains just the 15 basic elements, but refinements (identified as *terms*, *Dublin Core terms*, or *dcterms*) may be used as part of the system.[49] There are two broad classes of refinements now used:

- **Element Refinements**: qualifiers that sharpen the focus of an element, that is, they make the meaning of an element narrower or more specific
- **Encoding Schemes**: qualifiers that identify a controlled vocabulary or a data type that aids in the interpretation of an element value; these are referred to as *Vocabulary Encoding Schemes* and *Syntax Encoding Schemes* (or datatypes)

A list of refinements is provided in the dcterms namespace. Some are used to qualify one of the 15 elements in the DCMES, but some may be used alone or in place of an element. Table 7.3 shows the DCMES, with qualifiers that apply to each element.[50]

Table 7.3. Some Refinements Used with the Dublin Core.

DC Element	Refinements / Additional DC Terms	Encoding Schemes (Vocabulary or Syntax)
Contributor	n/a	n/a
Coverage	spatial	DCMI-Box; DCMI-Point; ISO 3166; TGN
	temporal	DCMI-Period; W3C-DTF
Creator	n/a	n/a
Date	available; created; dateAccepted; dateCopyrighted; dateSubmitted; issued; modified; valid	DCMI-Period; W3C-DTF
Description	abstract; tableOfContents	n/a
Format	extent; medium	Internet Media Types (IMT)
Identifier	bibliographicCitation	URI
Language	n/a	ISO 639-2; ISO 639-3; RFC 1766; RFC 3066; RFC 4646
Publisher	n/a	n/a
Relation	conformsTo; hasFormat; hasPart; hasVersion; isFormatOf; isPartOf; isReferencedBy; isReplacedBy; isRequiredBy; isVersionOf; references; replaces; requires	URI
Rights	accessRights; license	n/a
Source	n/a	URI
Subject	n/a	DDC; LCC; LCSH; MeSH; NLM; UDC
Title	alternative	n/a
Type	n/a	DCMIType
DC terms associated with educational materials	audience educationLevel instructionalMethod mediator	

DC terms associated with collections	**accrualMethod** **accrualPeriodicity** **accrualPolicy**
Additional DC terms	**provenance** **rightsHolder**

Dublin Core is used worldwide. The DCMI currently supports several "communities"[51] that bring together interest groups for people in a particular domain or who are engaged in specific topics within the DCMI framework. In addition, the standard is used to organize resources ranging from simple websites to complex government information. It has been used in catalogs, databases, and digital library applications. Among these, as noted already, is OCLC with an implementation in its WorldCat database. Catalogers using this system can view and download records encoded with MARC or encoded with an HTML or XML/RDF implementation of DC. A number of application profiles based on DC are in existence. An example is the *Library Application Profile*, which defines required elements, permitted qualifiers and dcterms, permitted encoding or vocabulary schemes, additional elements from other metadata schemas (such as MODS), and so on.[52]

Metadata Object Description Schema (MODS)

MODS was discussed in Chapter 6 as an encoding standard. It is actually a hybrid metadata schema that incorporates both encoding rules and a set of named elements. It was developed by the LC Network Development and MARC Standards Office in consultation with other experts.[53] MODS provides an alternative "between a very simple metadata format with a minimum of fields and no or little substructure (for example, DC) and a very detailed format with many data elements having various structural complexities such as MARC 21."[54] In other words, it is richer than DC but simpler than the MARC format. Its development as an XML standard means that its language-based tags can be understood by any English-speaking person; although as with all language-based tagging, international use by non-English readers could be problematic.

MODS was designed with a strong affinity to MARC. However, when MODS was designed, it was decided that some MARC elements would be combined into a single MODS element and some MARC fields would be dropped altogether. In addition some MARC elements recur in more than one element as sub-elements. For example, *name, identifier,* and *titleInfo* can be used as both elements and sub-elements with the same definition for each. *Name,* for example, can be the primary name associated with the resource, or it can be a name associated with a related information resource.[55]

MODS has top-level elements followed by sub-elements. The 20 top-level elements are:

- titleInfo
- name
- typeOfResource
- genre
- originInfo
- language
- physicalDescription
- abstract
- tableOfContents
- targetAudience
- note
- subject
- classification
- relatedItem
- identifier
- location

- accessCondition
- part
- extension
- recordInfo

An example of sub-elements follows, showing the sub-elements under *originInfo:*

- place
- publisher
- dateIssued
- dateCreated
- dateCaptured
- dateValid
- dateModified
- copyrightDate
- dateOther
- edition
- issuance
- frequency

MODS was made available officially in June 2002, and experimentation and implementation have proceeded since then. The current version of MODS is the MODS 3.6 schema (released in 2015). As with preceding versions, an accompanying list of approved changes from MODS 3.5 is provided.[56] The MODS Implementation Registry enumerates many projects that are using MODS.[57] There are nearly 40 projects listed at the time of this writing. For example, the Amherst College Digital Collections has been using MODS for descriptive metadata for all digitized materials. "This repository holds tens of thousands of digitized art works, photographs, manuscripts, books, and other media from the Archives & Special Collections and the library's Art & Architecture Collection."[58] (See Figure 7.5 for a sample MODS record.)

```xml
<mods xmlns="http://www.loc.gov/mods/v3" xmlns:xsi="http://www.w3.org/2001/XMLSchema-
instance" xsi:schemaLocation="http://www.loc.gov/mods/v3 http://www.loc.gov/standards/mods/v3/mods-3-
5.xsd" version="3.5">
        <titleInfo>
                <title>Introduction to cataloging and classification</title>
        </titleInfo>
        <name type="personal" usage="primary">
                <namePart>Joudrey, Daniel N.</namePart>
                <role><roleTerm type="text">author.</roleTerm></role>
        </name>
        <name type="personal">
                <namePart>Taylor, Arlene G.</namePart>
                <namePart type="date">1941-</namePart>
                <role><roleTerm type="text">author.</roleTerm></role>
        </name>
        <name type="personal">
                <namePart>Miller, David P. (David Peter)</namePart>
                <namePart type="date">1955-</namePart>
                <role><roleTerm type="text">author.</roleTerm></role>
        </name>
        <typeOfResource>text</typeOfResource>
        <genre authority="marcgt">bibliography</genre>
        <genre authority="rdacontent">text</genre>
        <originInfo>
                <place>
                        <placeTerm type="code" authority="marccountry">cau</placeTerm>
                </place>
                <dateIssued encoding="marc">2015</dateIssued>
                <edition>Eleventh edition.</edition>
                <issuance>monographic</issuance>
        </originInfo>
        <originInfo eventType="publication">
                <place>
                        <placeTerm type="text">Santa Barbara, California :</placeTerm>
                </place>
                <publisher>Libraries Unlimited,</publisher>
                <dateIssued>[2015]</dateIssued>
        </originInfo>
        <language>
                <languageTerm type="code" authority="iso639-2b">eng</languageTerm>
        </language>
        <physicalDescription>
                <form authority="marcform">print</form>
                <extent>xxv, 1048 pages : illustrations ; 27 cm.</extent>
                <form type="media" authority="rdamedia">unmediated</form>
                <form type="carrier" authority="rdacarrier">volume</form>
        </physicalDescription>
        <note type="statement of responsibility" altRepGroup="00">Daniel N. Joudrey, Arlene G. Taylor, and David P.
        Miller.</note>
        <note>Significant expansion of: Introduction to cataloging and classification. Tenth edition / Arlene G. Taylor ; with
        the assistance of David P. Miller.</note>
        <note type="bibliography">Includes bibliographical references (pages 1001-1022) and index.</note>
```

```xml
        <subject authority="lcsh">
                <titleInfo><title>Resource description & access</title></titleInfo>
        </subject>
        <subject authority="lcsh">
                <topic>Descriptive cataloging</topic>
        </subject>
        <subject authority="lcsh">
                <topic>Subject cataloging</topic>
        </subject>
        <subject authority="lcsh">
                <topic>Classification</topic>
                <topic>Books</topic>
        </subject>
        <subject authority="bisacsh">
                <topic>LANGUAGE ARTS & DISCIPLINES / Library & Information Science / Cataloging &
                Classification</topic>
        </subject>
        <subject authority="bisacsh">
                <topic>LANGUAGE ARTS & DISCIPLINES / Study & Teaching</topic>
        </subject>
        <classification authority="lcc">Z693 .W94 2015</classification>
        <classification authority="ddc" edition="23">025.3</classification>
        <classification authority="bisacsh">LAN025030 LAN020000</classification>
        <relatedItem type="series">
                <titleInfo><title>Library and information science text series</title>
                </titleInfo>
        </relatedItem>
        <relatedItem>
                <titleInfo><title>Introduction to cataloging and classification</title></titleInfo>
                <name type="personal">
                        <namePart>Taylor, Arlene G.,</namePart>
                        <namePart type="date">1941-</namePart>
                </name>
        </relatedItem>
        <identifier type="isbn">9781598848571 (hardback : acid-free paper)</identifier>
        <identifier type="isbn">9781598848564 (pbk : acid-free paper)</identifier>
        <identifier invalid="yes" type="isbn">9781440837456 (ebook)</identifier>
        <identifier type="lccn">2015012911</identifier>
        <recordInfo>
                <descriptionStandard>rda</descriptionStandard>
                <recordContentSource authority="marcorg">DLC</recordContentSource>
                <recordCreationDate encoding="marc">150616</recordCreationDate>
                <recordChangeDate encoding="iso8601">20160818094544.0</recordChangeDate>
                <recordIdentifier>18660797</recordIdentifier>
                <recordOrigin>Converted from MARCXML to MODS version 3.5 using MARC21slim2MODS3-5.xsl
                (Revision 1.106 2014/12/19)</recordOrigin>
                <languageOfCataloging>
                        <languageTerm type="code" authority="iso639-2b"> eng</languageTerm>
                </languageOfCataloging>
        </recordInfo>
</mods>
```


Figure 7.5. An Excerpt from a MODS Record. (MODS coding shown in bold)

Archives Metadata Standards

The archival community was relatively late in developing community metadata strategies. Tremendous progress, however, has been made over the past 30 years on the international and national fronts. A few metadata standards related to archives are discussed here as examples.

General International Standard Archival Description (ISAD(G))

ISAD(G)[59] is an international standard providing guidance for archival description. As with ISBD, the intention is that it be used as a basis to develop national standards. It is designed to facilitate the creation of archival descriptions that "identify and explain the context and content of archival material in order to promote its accessibility."[60] Although much information is acquired and recorded at every stage of the management of archival resources, the ISAD(G) rules are for description of materials starting at the point that it has been decided to preserve and control them. The rules in the standard do not explain how to describe special materials such as sound recordings, maps, and so on. Manuals that already exist for such materials are expected to be consulted as needed.

There are 26 elements identified in ISAD(G). Each has rules that include the name of the element, a statement of the purpose of the element in a description, the rules that apply to the element, and examples, if possible. The elements are organized into seven areas of descriptive information:

- **Identity Statement Area**: information essential to identify the unit being described
- **Context Area**: information about the origin and custody of the unit being described
- **Content and Structure Area**: information about the subject matter and the arrangement of the unit being described
- **Condition of Access and Use Area**: information about the availability of the unit being described
- **Allied Materials Area**: information about other information resources that have an important relationship to the unit being described
- **Note Area**: information that cannot be accommodated in any of the other areas
- **Description Control Area**: information about the preparation of the archival description (e.g., by whom, when, how, when revised)

Six elements are considered essential for international exchange of descriptive information:

- **Reference code**: code made up of standardized country code, repository code, and local repository specific numbers
- **Title**: concise title conveying authorship, subject matter, and form of material
- **Creator**: originator of a particular collection
- **Date(s)**: dates of creation or subject matter, depending upon the nature of the unit being described
- **Extent of the unit of description**: statement of the bulk, quantity, or size
- **Level of description**: statement of the grouping being described (e.g., fonds or sub-fonds; series, subseries, file, or item)

Five of the six core elements are derived from the Identity Area, with only the Creator element coming from the Context Area.

ISAD(G) is based on four principles:

- Description proceeds from the general to the specific.
- Information should be relevant to the level of description.
- Descriptions should be linked between levels.
- Information should not be repeated.

These principles subsequently are found in many archival descriptive standards that are considered to be ISAD(G) compliant (e.g., DACS in the United States, *Rules for Archival Description* [RAD] in Canada) or are used directly as the basis for archival description (e.g., the Archives Hub in the United Kingdom).[61] The standard does not dictate the encoding that is to be used, although there has been much cooperation with the development of the EAD standard (see below).

Describing Archives: A Content Standard (DACS)

DACS[62] is a standard for the description of archival materials that has been accepted by most of the U.S. archival community. It supersedes *Archives, Personal Papers, and Manuscripts* (APPM),[63] a content standard developed by Steve Hensen while at LC that was based on AACR2 and was used by the archival community as a standard for the creation of MARC catalog records. (AACR2 has only a skeletal chapter for manuscripts; it is geared toward illuminated manuscripts and itself acknowledges the need for other content standards to be used for archival descriptions.) DACS is based on the international standard for archives, ISAD(G), and can be used to create any type or level of description of archival and manuscript materials, including catalog records and full finding aids.[64]

The second edition of DACS has two parts—one for describing archival materials and one for archival authority records. The second part is discussed in Chapter 8 of this text. Among its prefatory materials, DACS includes a statement of principles that restate generally accepted archival theory and constitute the basis for the standard. These principles are

- **Principle 1**: Records in archives possess unique characteristics.
- **Principle 2**: The principle of *respect des fonds* is the basis of archival arrangement and description.
- **Principle 3**: Arrangement involves the identification of groupings within the material.
- **Principle 4**: Description reflects arrangement.
- **Principle 5**: The rules of description apply to all archival materials, regardless of form or medium.
- **Principle 6**: The principles of archival description apply equally to records created by corporate bodies, individuals, or families.
- **Principle 7**: Archival descriptions may be presented at varying levels of detail to produce a variety of outputs.
    - **Principle 7.1**: Levels of description correspond to levels of arrangement.
    - **Principle 7.2**: Relationships between levels of description must be clearly indicated.
    - **Principle 7.3**: Information provided at each level of description must be appropriate to that level.
- **Principle 8**: The creators of archival materials, as well as the materials themselves, must be described.[65]

In 2016 work began on reviewing and revising these principles so that they will more adequately reflect the values and objectives of the archival community. An updated list of principles is expected in 2017 or 2018.[66]

The first chapter of DACS discusses levels of description, including guidelines for single-level description (only one level of description) and multilevel description (one level of description with *at least one* subordinate level of description). Like other content standards, it also contains guidelines for minimum, optimum, and added value descriptions, based on the number of elements included in the description. Chapters 2–8 present rules for the 25 elements of archival description.

2. **Identity Elements**

    2.1 Reference Code (Required)
    2.2 Name and Location of Repository (Required)
    2.3 Title (Required)
    2.4 Date (Required)
    2.5 Extent (Required)
    2.6 Name of Creator(s) (Required, if known)
    2.7 Administrative/Biographical History (Optimum)

3. **Content and Structure Elements**
   - 3.1 Scope and Content (Required)
   - 3.2 System of Arrangement (Added value)
4. **Conditions of Access and Use Elements**
   - 4.1 Conditions Governing Access (Required)
   - 4.2 Physical Access (Added value)
   - 4.3 Technical Access (Added value)
   - 4.4 Conditions Governing Reproduction and Use (Added value)
   - 4.5 Languages and Scripts of the Material (Required)
   - 4.6 Finding Aids (Added value)
5. **Acquisition and Appraisal Elements**
   - 5.1 Custodial History (Added value)
   - 5.2 Immediate Source of Acquisition (Added value)
   - 5.3 Appraisal, Destruction, & Scheduling Infor-ma-tion (Added value)
   - 5.4 Accruals—anticipated additions to the unit being described (Added value)
6. **Related Materials Elements**
   - 6.1 Existence and Location of Originals (Added value)
   - 6.2 Existence and Location of Copies (Added value)
   - 6.3 Related Archival Materials (Added value)
   - 6.4 Publication Note (Added value)
7. **Notes Elements**
   - 7.1 Notes (Added value)
8. **Description Control Elements**
   - 8.1 Description Control (Added value)

There are some major differences between describing archival materials and describing library resources. Aside from focusing on aggregate descriptions, archival materials usually do not have established or given titles, so the archivist must assign one. Guidelines are given for creating archivist-supplied titles (i.e., *devised titles*). Additionally, archives generally contain a wide variety of different material types, including for example, materials that could be handwritten, correspondence, photographs, newspaper clippings, reports, and so on. Another difference is the application of the term *series* in archives. Archival materials do not have series as defined in the library community; archival series are another concept entirely. In archival collections, a *series* is a logical group of files, books, correspondence, or other sets of documents and constitutes a basic building block in multilevel descriptions. This concept is discussed in Chapter 1 of this text. Although DACS is a schema-neutral content standard that does not prescribe any particular method of output, it provides examples that could be encoded with EAD or with MARC 21.

Encoded Archival Description (EAD)

The *EAD Version 2002,*[67] discussed as an XML schema in Chapter 6, was created specifically to encode finding aids. Its design principles state: "EAD is a data structure and not a data content standard. It does not prescribe how one formulates the data that appears in any given data element—that is the role of external national or international data content standards."[68] Guidelines for finding aids are already in existence (e.g., ISAD(G) and DACS discussed above), so EAD does not need to contain prescriptions for content. However, it has a header section that was originally based on the TEI Header (see below). The header contains metadata about the creation and maintenance of the EAD document itself. In many instances archival collections have been cataloged using MARC instead of, or in addition to, EAD finding aids being created.

EAD3 was released in 2015, and at the time of this writing is starting to be adopted by the archival community: "EAD3 seeks to simplify EAD and to update EAD to connect more easily with other standards like EAC-CPF."[69] Changes to the schema represented in EAD3 are outlined in the Preface to the EAD3 Tag Library and are driven by four principles:

- "Achieving greater conceptual and semantic consistency in the use of EAD."
- "Exploring mechanisms whereby EAD-encoded information might more seamlessly and effectively connect with, exchange, or incorporate data maintained according to other protocols."
- "Improving the functionality of EAD for representing descriptive information created in international and particularly in multilingual environments."
- "Being mindful that a new version will affect current users."[70]

For a complete listing of elements used in EAD3, please see the latest version of the tag library at the EAD Official Site.[71] For an example of an EAD3-encoded finding aid, see Appendix D in this text.

Other Domain-Specific Metadata Schemas

Many metadata schemas have been developed by different communities to be used in specific situations for specialized resources. They support description that allows for details needed by users searching for resources in a particular domain. A few are discussed here as examples.

TEI (Text Encoding Initiative) Headers

As discussed in Chapter 6, TEI was originally an SGML DTD (now available as an XML schema) that was created to provide a way of encoding literary and/or scholarly texts so that encoded versions could be exchanged easily. TEI represents a constellation of encoding protocols for various literary forms. It is suggested in guidelines that one needs to customize TEI to get the best from it. A specific popular customization that is used by a core constituency is *TEI Lite*.[72]

One part of TEI is the TEI Header, created so that there is metadata for the electronic version of the text as part of the text file. A strong motivation for the creation of a standard for the TEI Header was to provide a source of information for cataloging. TEI creators collaborated with library catalogers for the *file description* <fileDesc> element of the header. They wanted the mapping from TEI to a catalog record and vice versa to be as simple as possible. The standard TEI Header requires many of the same elements as does ISBD; therefore, the header content greatly resembles library cataloging.

The TEI Header has five sections, only one of which is required. The five components are listed below:

- **File Description <fileDesc>**: The *file description* is required and contains a bibliographic description of the electronic version of the text. It includes the title statement, with author information; edition statement; extent; publication statement; series statement; notes statement; and the source description, which is a description of the original source from which the electronic text was derived.
- **Encoding Description <encodingDesc>**: This section explains the relationship between an electronic text and its source, and what rules were used or editorial decisions were made in transcribing the text (e.g., information about how quotations and spelling variations were treated, information about the tags used in the document). This area includes elements that cover project description, tagging, text sampling, references, geographic coordinates, and so on.
- **Text Profile <profileDesc>**: The *text profile* has information characterizing various nonbibliographic descriptive aspects of a text. It contains language information, subject terminology, and classification notation(s) as well as contextual information to help understand the situation in which the resource was produced. This section may include abstracts, creation information, language usage, classifications, correspondence descriptions, and a description of the calendar system used for dates in the resource.
- **Container Element <xenoData>**: This section contains any non-TEI metadata generated that the project wishes to include.
- **Revision History <revisionDesc>**: The *revision history* contains a record of every change that has been made to the electronic version of the text, including when each change was made and by

whom.

The form of the content that is entered into the fields is not dictated in the guidelines for TEI Headers. For an example of a TEI Header, refer back to Figure 6.8 on page 273.

Cataloging Cultural Objects (CCO)

Cataloging Cultural Objects (CCO) is a content standard like RDA and DACS, but one that has been designed for communities that describe works of art, architecture, cultural artifacts, and images of these things.[73] CCO instructs catalogers to create separate records for works and for images of those works. Although the VRA Core and CDWA (discussed below) provide lists and descriptions of elements or categories for metadata element sets, CCO prescribes the data values and defines the order, syntax, and form in which the values are to be entered into a data structure. Similar to DACS, CCO concentrates on principles of good cataloging and documentation, not on rigid rules. Catalogers are expected to make informed judgments. The following are the 10 key principles of CCO:

1. Establish the logical focus of each Work Record, whether it is a single item, a work made up of several parts, or a physical group or collection of works. Clearly distinguish between Work Records and Image Records.
2. Include all the required CCO elements.
3. Follow the CCO rules. Make and enforce additional local rules to allow information to be retrieved, repurposed, and exchanged effectively.
4. Use controlled vocabularies, such as the Getty vocabularies and Library of Congress authorities.
5. Create local authorities that are populated with terminology from standard published controlled vocabularies as well as with local terms and names. Structure local authorities as thesauri whenever possible. Record and document decisions about local authorities.
6. Use established metadata standards, such as the *VRA Core Categories* or *Categories for the Description of Works of Art*.
7. Understand that cataloging, classification, indexing, and display are different but related functions.
8. Be consistent in establishing relationships between works and images, between a group or collection and works, among works, and among images.
9. Be consistent regarding capitalization, punctuation, and syntax. Avoid abbreviations, but when necessary, use standard codes and lists for abbreviations (for example, the ISO abbreviations for countries).
10. For English-language information systems and users, use English-language data values whenever possible.[74]

CCO is divided into three parts. Part One, *General Guidelines,* deals with such issues as what unit is being cataloged, what is a work and what is an image, the relationships between works and images, specificity and exhaustivity of cataloging, and the different kinds of relationships that may be delineated. Although CCO does not discuss administrative and technical metadata, Part One includes a discussion of database design issues and what may be needed for such designs. Lastly, Part One discusses authority files and controlled vocabularies.

Part Two of CCO, *Elements,* is divided into chapters on: object naming; creator information; physical characteristics; stylistic, cultural, and chronological information; location and geography; subject; class; description; and view information. Each chapter contains discussion of the concept, the rules for the elements covered by that chapter, and ways for presenting the data for display and indexing. Some elements are marked as required. These are expected to appear in the description as long as the element is applicable and discernible. These include the following:

- **Object Naming**
  - Work Type
  - Title

- **Creator Information**
  - Creator display
  - Controlled Creator
- **Physical Characteristics**
  - Measurements display
  - Materials and Techniques display
- **Stylistic, Cultural, and Chronological Information**
  - Display Date
  - Earliest Date
  - Latest Date
- **Location and Geography**
  - Current Location display
- **Subject**
  - Controlled Subject
- **View Information**
  - View Description
  - View Type
  - View Subject Controlled

Additional required elements are listed for Part Three of CCO, which covers authorities for personal and corporate names, geographic places, concepts, and subjects. These are discussed further in Chapter 8 of this text.

VRA (Visual Resources Association) Core

According to the VRA Core website, the element set provides a categorical organization for "the description of works of visual culture as well as the images that document them."[75] Version 4.0 of the VRA Core was released in 2007 and last updated in 2014.[76] There are three main supporting documents (in English) in this version: an introduction, an element outline, and the element descriptions with tagging examples. (Documentation is also available in Italian and Greek.)

The element description document includes the names of the categories, each followed by a definition, information about attributes and sub-elements, recommendations for use of controlled vocabularies or standardized lists, whether or not the element is required and repeatable, and mappings to the previous version of the VRA Core Categories, the *Categories for the Description of Works of Art* (CDWA) (see below), CCO, and DC elements. A VRA Core 4.0 XML Schema was developed along with Core 4.0, and in the document of element descriptions examples are given encoded with XML. There are 19 basic categories of elements:

- **Work, Collection, or Image**: identifies the record as being for a *work* (a physical or created object or an event), for an *image* (a visual surrogate of a work), or for a *collection* (an aggregate of works or images)
- **Agent**: identifies the names of entities that have contributed to the design, creation, or production of the work or image (sub-elements for *name*, *culture*, *role*, *dates*, and *attribution*; use of authority files recommended)
- **Cultural Context**: identifies a culture, people, or country from which a work, image, or collection originates or with which it has been associated (use of controlled terms recommended)
- **Date**: identifies date(s) associated with appearance of the work or image (the type of date attribute may be creation, design, production, alteration, restoration, and so on, and sub-elements are for *earliest date* and *latest date*; use of standard date presentations recommended)

- **Description**: a free-text note that may include comments, description, interpretation, and any other information not recorded in other elements
- **Inscription**: identifies marks or written words (e.g., signature, date, dedication) added to the object either at its production or later in its history (sub-elements may identify *author*, *position*, and *type of text*)
- **Location**: identifies geographic location and/or a specific site location of the work or image (sub-elements include *name* and *refid* [i.e., an identifying number or code]; use of controlled terms recommended)
- **Material**: identifies the substance that composes a work or image; often the material(s) used for the work (e.g., paper and ink, bronze; use of *Art & Architecture Thesaurus* (AAT) terms recommended)[77]
- **Measurements**: identifies the size, shape, scale, dimensions, or format of the work or image (attributes include types of dimensions and types of units)
- **Relation**: identifies a related work or image and the relationship that exists between the described and the related works or images (includes a lengthy list of recommended relationships for the relationship type attribute, e.g., partOf, modelFor, mateOf, copyAfter)
- **Rights**: identifies copyright information, intellectual property statements, or other information needed for rights management (sub-elements include *rights holder* and *text* [of the rights statement])
- **Source**: identifies the source of the information recorded about the work or image (e.g., a bibliographic citation to a book that describes the work or image; sub-elements include *name* and *refid* [i.e., an identifying number or code])
- **State Edition**: identifies the number or name of a state, edition, or impression of a work that exists in more than one form (sub-elements include *name* and *description*; attributes identify the number of known states, editions, or impressions and which one this is)
- **Style Period**: identifies a defined style, historical period, group, movement, etc., characteristics of which appear in the work or image (use of AAT terms recommended)
- **Subject**: gives terms or phrases that describe or interpret the work or image and that represent what it depicts or expresses (use of controlled terms recommended)
- **Technique**: identifies the processes, techniques, or methods used in making or altering the work or image (use of *AAT* terms recommended)
- **Textref**: gives identification values derived from textual references (e.g., catalog number) that are not associated with a particular repository or other location (sub-elements include *name* and *refid* [i.e., an identifying number or code])
- **Title**: the identifying phrase given to the work or image (use of data content rules for titles of works of art recommended)
- **Work Type**: identifies the specific type of work, image, or collection being described (use of *AAT* terms recommended)

Guidelines for description using the VRA Core are included in CCO (described above).[78]

Categories for the Description of Works of Art (CDWA)

Categories for the Description of Works of Art (CDWA) is another metadata element set that, like the VRA Core, is used for describing and accessing information about works of art, architecture, other material culture, and related images.[79] CDWA comprises 512 categories and sub-categories. Of these, 36 are considered *core*—that is, the ones required to describe a work in a unique and unambiguous way. The CDWA core categories are divided into six groupings based on their purpose or focus:[80]

- For the object, architecture, or group
- For related textual references authority

- For creator identification authority
- For place/location authority
- For generic concept authority
- For subject authority

The core categories under the first group are:

- Catalog Level
- Object/Work Type
- Classification Term
- Title or Name
- Measurements Description
- Materials and Techniques Description
- Creator Description
- Creator Identity
- Creator Role
- Creation Date
    - Earliest Date
    - Latest Date
- Subject Matter Indexing Terms
- Current Location Repository Name/Geographic Location
- Current Repository Numbers

The remaining core categories are for access and authority control and are discussed in Chapter 8.

CDWA, like VRA Core, recommends encoding records with XML and has developed an XML schema called *CDWA Lite* containing a smaller set of essential elements extracted from the fuller version of CDWA.[81] It comprises 22 core elements, most of which have between one and nine sub-elements. CDWA Lite records are intended to be contributions to union catalogs and other repositories through harvesting protocols such as the Open Archives Initiative (OAI).

ONIX (Online Information Exchange)

The publishing world has been developing standards that can be used for description of resources by their publishers. ONIX (Online Information Exchange) for Books,[82] now available in Release 3.0.3, was discussed in the encoding standards chapter as an XML DTD and schema. It is maintained by EDItEUR jointly with several book industry and user groups. It includes a data dictionary that defines the content of the elements. The elements allow a publisher to present online the information that was previously contained on book jackets and in publishing brochures and catalogs—information such as synopses, quotations from reviews, author biographies, intended audience, and so on. As mentioned in Chapter 6, ONIX uses XML to transmit ONIX messages. An ONIX message contains both (1) a header that contains information about the sender, addressee, and administrative metadata, and (2) a product record that contains information about the resource being described. The product record may contain six blocks of data:

- Product description
- Marketing collateral detail
- Content detail
- Publishing detail
- Related material

- Product supply

The product description block contains information most similar to traditional library cataloging data found in RDA, AACR2, or ISBD. A mapping to MARC allows consideration of using ONIX data in library catalogs.[83] There are also projects that link MARC records in catalogs to ONIX data held in XML-encoded databases.

A partnership between EDItEUR and NISO has been responsible for the production of ONIX for Serials, a family of XML formats for information about serial products and subscription information.[84] This has led to the development of additional ONIX formats, such as SPS (Serials Products and Subscriptions), SOH (Serials Online Holdings), and SRN (Serials Release Notification). The ONIX Serials Coverage Statement is an XML structure for use in the three formats. An example of an ONIX message is shown in Figure 6.9.

Index and Bibliography Records

At the moment, there is no official standard for the descriptive content of index records. An ANSI standard Z39.4–1984, subtitled "Basic Criteria for Indexes,"[85] has no mention of the data to be included in the description of an item. The National Information Standards Organization (NISO) tried to update this standard, but committees could not come to agreement with the American Society for Indexing, and so a proposed revision of Z39.4 was withdrawn in 1996. In 1997 the rejected replacement for Z39.4 was instead released as a NISO technical report: TR-02, *Guidelines for Indexes and Related Information Retrieval Devices*. Although its author, James Anderson, presented useful, updated information in TR-02 to help fill the void, it was not an official NISO standard.[86]

The second edition of ISO 999 *Information and Documentation—Guidelines for the Content, Organization and Presentation of Indexes* was released as an international standard in 1996.[87] It was intended to update, clarify, and exemplify basic indexing methods that were presented in the original 1975 edition. It contains additional information on the organization of indexes, quality control, and arrangement. It was reviewed and reconfirmed in 2015. The standard contains sections that address the following areas:

- Scope
- Associated standards
- Definitions
- Functions
- Types of indexes (Subject, Author, Name, Geographic, Title, and Number/Code indexes)
- Quality control (Length, Detail, Consistency, etc.)
- Content and organization (Construction, Concepts, Proper names, Locators, References, etc.)
- Arrangement (Filing order, Word-by-word versus Letter-by-letter, Alphanumeric arrangement, Subheadings, etc.)
- Presentation (Notes, Search aids, Style, etc.)

Creators of indexes each have their own standards for information to be included in an index record. Agencies like H. W. Wilson that publish several different indexes in different subject areas have some consistency from index to index. Electronic index records tend to include more information than do paper versions. As in OPACs, electronic index records tend to have labels, which are not included in paper versions. (For an example of a labeled index record, refer back to Figure 3.9 on page 120.)

There is one standard for bibliographic references, ISO 690:2010[88] guiding the content that is to be included in entries in bibliographies. It developed as an ISO Technical Report and replaced two earlier standards with more up-to-date guidance on creating bibliographic references and citations for a wide range of information resources, both tangible and electronic.

CONCLUSION

From the preceding sections, one can see that the creation of resource descriptions is somewhat dependent upon the community for which the records are being created. Some communities have long histories of using standards. The library community has the longest traditions, being based upon principles that have been developing for centuries. Other communities have recognized the library's long experience and have patterned their guidelines for creation of surrogate records after the ISBD or other standard. The creators of TEI Headers, DC, VRA Core, CDWA, CCO, and others have included librarians, along with other people with organizing experience, in their planning committees.

Some Important Terms in This Chapter

Access Point	Edition	Other Title Information
Accompanying Material	Element	Parallel Title
Agent	Element Refinement	Policy Statement
Alternative Title	Exceptions	Preferred Sources of Information
Alternatives	Expression	Qualifier
Analytical Description	Four-Fold Path	RDA Toolkit
Attribute	General Material Designation	Resource
Bibliographic Record	Hierarchical Description	Rule of Three
Cataloging Code	Information Resource	Schema-Neutral
Chief Source of Information	Integrating Resource	Serial
Collection-Level Description	Interoperability	Series
Collective Title	Item	Single Unit
Comprehensive Description	Manifestation	Statement of Responsibility
Content Standard	Metadata	Subtitle
Contributor	Metadata Statement	Surrogate Record
Conventional Collective Title	Mode of Issuance	Syntax
Core Element	Monograph	Title
Core-If Element	Multipart Monograph	Title Proper
Creator	One-to-One Principle	Triplestore
Descriptive Metadata	Optional Addition	User Tasks
Devised Title	Optional Omission	Value

This chapter has addressed the descriptive part of metadata, which we have called *resource descriptions.* In discussion of the value of resource descriptions, the chapter concludes that they are most helpful when they are predictable in both form and content. The most current standards in the United States to prescribe form of content are RDA, DACS, and CCO. The remaining standards discussed here rarely prescribe form of content, although they prescribe and define elements. Perhaps experience with the large quantities of metadata records that have long existed in the library and archival communities, and more recently in the art and museum communities, will lead some creators of metadata to re-evaluate their need for prescribed content. The next chapter addresses access and authority control, which is accepted as a necessity in the library and archival communities—and is highly recommended in the art and museum communities—but has barely been considered in the other communities.

Some Important Acronyms in This Chapter

AACR:	*Anglo-American Cataloging Rules*
AACR2:	*Anglo-American Cataloguing Rules, Second edition*
ANSI:	American National Standards Institute
ASIS&T:	Association for Information Science and Technology
CCO:	*Cataloging Cultural Objects*
CDWA:	*Categories for the Description of Works of Art*
DACS:	*Describing Archives: A Content Standard*
DC:	Dublin Core
DCMES:	Dublin Core Metadata Element Set
DCMI:	Dublin Core Metadata Initiative
DOI:	Digital Object Identifier
DTD:	Document Type Definition
EAD:	Encoded Archival Description
FRAD:	*Functional Requirements for Authority Data*
FRBR:	*Functional Requirements for Bibliographic Records*
HTML:	Hypertext Markup Language
ICP:	*Statement of International Cataloguing Principles*
IFLA:	International Federation of Library Associations and Institutions
IRI:	Internationalized Resource Identifier
ISAD(G):	*General International Standard Archival Description*
ISBD:	*International Standard Bibliographic Description*
ISBN:	International Standard Book Number
ISO:	International Organization for Standardization
LC:	Library of Congress
LC-PCC PS:	*Library of Congress-Program for Cooperative Cataloging Policy Statements*
LCRI:	*Library of Congress Rule Interpretations*
LRM:	*IFLA Library Reference Model*
MODS:	Metadata Object Description Schema
NISO:	National Information Standards Organization
ONIX:	Online Information Exchange
RDA:	*Resource Description & Access*
RDF:	Resource Description Framework
TEI:	Text Encoding Initiative
URI:	Uniform Resource Identifier
URL:	Uniform Resource Locator
VRA:	Visual Resources Association
XML:	Extensible Markup Language

NOTES

All URLs accessed June 2017.

1. Daniel N. Joudrey, Arlene G. Taylor, and David P. Miller, *Introduction to Cataloging and Classification*, 11th ed. (Santa Barbara, CA: Libraries Unlimited, 2015), 161.

2. Murtha Baca, Patricia Harpring, Elisa Lanzi, Linda McRae, and Ann Whiteside, on behalf of the Visual Resources Association, *Cataloging Cultural Objects*: *A Guide to Describing Cultural Works and Their Images* (Chicago: American Library Association, 2006), 3, http://cco.vrafoundation.org/index.php/toolkit/cco_pdf_version/ [cited hereinafter as CCO].

3. *RDA: Resource Description & Access*, developed in a collaborative process led by the Joint Steering Committee for Development of RDA (Chicago: American Library Association, 2010). Also available in *RDA Toolkit: Resource Description & Access* (Chicago: American Library Association, 2010), http://www.rdatoolkit.org/. In lieu of page numbers, any references to specific parts of RDA are made using instruction numbers.

4. International Federation of Library Associations and Institutions, IFLA Study Group, *Functional Requirements for Bibliographic Records: Final Report* (Munich: Saur, 1998), https://www.ifla.org/publications/functional-requirements-for-bibliographic-records. In following endnotes, this resource is identified as "FRBR."

5. Pat Riva, Patrick Le Boeuf, and Maja Žumer, "IFLA Library Reference Model: A Conceptual Model for Bibliographic Information," August 2017 (The Hague: IFLA, 2017), https://www.ifla.org/publications/node/11412.

6. CCO, 48.

7. *Describing Archives: A Content Standard* (DACS), 2nd ed. (Chicago: Society of American Archivists, 2013), 21–22. Also, freely available on the web at: http://www2.archivists.org/groups/technical-subcommittee-on-describing-archives-a-content-standard-dacs/dacs.

8. IFLA, FRBR, https://www.ifla.org/publications/functional-requirements-for-bibliographic-records.

9. RDA 1.5.3 Analytical Description.

10. Heather Dunn, "Collection Level Description: The Museum Perspective," *D-Lib Magazine* 6, no. 9 (2000), http://www.dlib.org/dlib/september00/dunn/09dunn.html.

11. RDA 2.2.2 Preferred Sources of Information.

12. DACS, 3.

13. CCO, 248.

14. Ibid., 50.

15. IFLA, Meetings of Experts on an International Cataloguing Code, *Statement of International Cataloguing Principles*, 2016 edition (The Hague: IFLA, 2016), http://www.ifla.org/files/assets/cataloguing/icp/icp_2016-en.pdf.

16. Library of Congress, "Library of Congress Announces Its Long-Range RDA Training Plan," March 2, 2012, https://www.loc.gov/catdir/cpso/news_rda_implementation_date.html.

17. European RDA Interest Group [EURIG] 2016 Annual Meeting, "EURIG Minutes of the Members' Meeting, 2016," http://www.slainte.org.uk/eurig/docs/EURIG2016/2016_EURIG_Minutes_rev.pdf; Alan Danskin, "EURIG Annual Report: 2016–2017: Presented on 8 May 2017 at University of Florence," http://www.casalini.it/eurig2017/presentations/danskin.pdf.

18. Andrew Osborn, "The Crisis in Cataloging," in *Foundations of Cataloging: A Sourcebook,* eds. Michael Carpenter and Elaine Svenonius (Littleton, CO: Libraries Unlimited, 1985), 90–103. Also available in *Library Quarterly* 11, no. 4 (October 1941): 393–411.

19. IFLA, FRBR, https://www.ifla.org/publications/functional-requirements-for-bibliographic-records; IFLA Working Group on Functional Requirements and Numbering of Authority Records (FRANAR), *Functional Requirements for Authority Data: A Conceptual Model*, approved by the Standing Committees of the IFLA Cataloguing Section and IFLA Classification and Indexing Section, 2009, amended and corrected through July 2013, http://www.ifla.org/publications/functional-requirements-for-authority-data.

20. Riva, Le Boeuf, and Žumer, "IFLA Library Reference Model," https://www.ifla.org/publications/node/11412

21. IFLA, ICP, http://www.ifla.org/files/assets/cataloguing/icp/icp_2016-en.pdf.

22. Program for Cooperative Cataloging (PCC) ISBD and MARC Task Group, "Revised Final Report," 2016, http://www.loc.gov/aba/pcc/documents/isbdmarc2016.pdf.

23. Riva, Le Boeuf, and Žumer, "IFLA Library Reference Model," https://www.ifla.org/publications/node/11412.

24. Joudrey, Taylor, and Miller, *Introduction to Cataloging and Classification*, Chapters 4–9.

25. *International Standard Bibliographic Description* (ISBD), recommended by the ISBD Review Group, approved by the Standing Committee of the IFLA Cataloguing Section, Consolidated ed. (Berlin: De Gruyter Saur, 2011), http://www.ifla.org/files/assets/cataloguing/isbd/isbd-cons_20110321.pdf.

26. International Federation of Library Associations and Institutions, "Superseded ISBDs," http://www.ifla.org/isbd-rg/superseded-isbd-s.

27. International Federation of Library Associations and Institutions, "ISBD Review Group," http://www.ifla.org/isbd-rg.

28. Joudrey, Taylor, and Miller, *Introduction to Cataloging and Classification*, 887.

29. PCC ISBD and MARC Task Group, "Revised Final Report," http://www.loc.gov/aba/pcc/documents/isbdmarc2016.pdf.

30. ISBD Consolidated ed., 41.

31. Ibid., 45.

32. IFLA, "ISBD Review Group," http://www.ifla.org/isbd-rg.

33. Joudrey, Taylor, and Miller, *Introduction to Cataloging and Classification*, Chapter 23.

34. *Anglo-American Cataloguing Rules, Second Edition, 2002 Revision* (AACR2R), prepared under the direction of the Joint Steering Committee for Revision of AACR (Ottawa: Canadian Library Association; Chicago: American Library Association, 2002, loose-leaf, with updates), D-2. Also available in Cataloger's Desktop, by subscription.

35. *Library of Congress Rule Interpretations,* 2nd ed. (Washington, DC: Cataloging Distribution Service, Library of Congress, 1990–2010). Also available on Cataloger's Desktop by subscription.

36. Arlene G. Taylor, *Introduction to Cataloging and Classification*, 10th ed., with the assistance of David P. Miller (Westport, CT: Libraries Unlimited, 2006).

37. Arlene G. Taylor and Daniel N. Joudrey, *The Organization of Information*, 3rd ed. (Westport, CT: Libraries Unlimited, 2009).

38. Dublin Core Metadata Initiative, "DCMI Home: Dublin Core," http://dublincore.org/.

39. Dublin Core Metadata Initiative (DCMI), "History of the Dublin Core Metadata Initiative," http://dublincore.org/about/history/.

40. Dublin Core Metadata Initiative (DCMI), "Dublin Core Metadata Element Set," Version 1.1, Issued 2012, http://dublincore.org/documents/dces/.

41. *The Dublin Core Metadata Element Set,* ANSI/NISO Z39.85–2012 (Baltimore, MD: NISO, 2013), http://www.niso.org/apps/group_public/download.php/10256/Z39-85-2012_dublin_core.pdf.

42. *Information and Documentation—The Dublin Core Metadata Element Set—Part 1: Core Elements,* ISO 15836:2017 (Geneva: ISO, 2017). Available for purchase at: https://www.iso.org/standard/71339.html.

43. "ASIS&T: Association for Information Science and Technology," https://www.asist.org/.

44. DCMI, "Expressing Dublin Core Metadata using the Resource Description Framework (RDF)," 2008, http://dublincore.org/documents/dc-rdf/, and "Expressing Dublin Core Description Sets Using XML (DC-DS-XML)," http://dublincore.org/documents/dc-ds-xml/.

45. Misha Wolf and Charles Wicksteed, "Date and Time Formats," W3C, http://www.w3.org/TR/NOTE-datetime.

46. Internet Assigned Numbers Authority, "Media Types," http://www.iana.org/assignments/media-types/media-types.xhtml.

47. "RFC4646: Tags for Identifying Languages," https://www.ietf.org/rfc/rfc4646.txt or http://www.faqs.org/rfcs/rfc4646.html.

48. See list of "DCMI Type Vocabulary" at: DCMI, "DCMI Metadata Terms," http://dublincore.org/documents/dcmi-terms/.

49. DCMI, "DCMI Metadata Terms," http://dublincore.org/documents/dcmi-terms/.

50. Ibid.

51. DCMI, "Dublin Core Work Structure: Active DCMI Communities," http://www.dublincore.org/groups/#communities.

52. DCMI-Libraries Working Group, "Library Application Profile," Dublin Core Metadata Initiative, http://dublincore.org/documents/library-application-profile/.

53. Library of Congress, Network Development and MARC Standards Office (NDMSO), "MODS: Metadata Object Description Schema," http://www.loc.gov/standards/mods/.

54. Rebecca S. Guenther, "MODS: The Metadata Object Description Schema," *Portal: Libraries and the Academy* 3, no. 1 (2003): 139.

55. Ibid., 140.

56. Library of Congress NDMSO, "Approved Changes for MODS 3.6," last updated March 1, 2016, http://www.loc.gov/standards/mods/changes-3-6.html.

57. Library of Congress NDMSO, "MODS Implementation Registry," http://www.loc.gov/standards/mods/registry.php.

58. Amherst College Library, "Amherst College Digital Collections," https://acdc.amherst.edu/about/acdc.

59. International Council on Archives, *ISAD(G): General International Standard Archival Description,* 2nd ed. (Ottawa: International Council on Archives, 2000), https://www.ica.org/sites/default/files/CBPS_2000_Guidelines_ISAD(G)_Second-edition_EN.pdf.

60. Ibid., 7.

61. JISC, "Archives Hub," https://archiveshub.jisc.ac.uk/isadg/.

62. DACS.

63. Steven L. Hensen, comp., *Archives, Personal Papers, and Manuscripts: A Cataloging Manual for Archival Repositories, Historical Societies, and Manuscript Libraries,* 2nd ed. (Chicago: Society of American Archivists, 1989).

64. DACS, vii.

65. DACS, xv–xix.

66. Maureen Callahan and Adrien Hilton, "Report: DACS Principles Meeting in March 2017," Society of American Archivists Council Meeting, May 16–17, 2017, Chicago, Illinois, Agenda Item VI.I, http://www2.archivists.org/sites/all/files/0517-VI-I-DACS-PrinciplesDevelop.pdf.

67. "EAD: Encoded Archival Description," Library of Congress, http://www.loc.gov/ead/.

68. "Design Principles for Enhancement to EAD (December 2002)," Library of Congress, http://www.loc.gov/ead/eaddesgn.html.

69. Society of American Archivists (SAA), "Frequently Asked Questions about EAD and EAD3," http://www2.archivists.org/groups/encoded-archival-description-ead-roundtable/frequently-asked-questions-about-ead-and-ead3.

70. SAA, *Encoded Archival Description Tag Library, Version EAD3*, Library of Congress, https://www.loc.gov/ead/EAD3taglib/index.html.

71. EAD.

72. Text Encoding Initiative, "TEI Lite," http://www.tei-c.org/Guidelines/Customization/Lite/index.xml.

73. CCO, 4.

74. Ibid., 2–3.

75. Library of Congress, "An Introduction to VRA Core," http://www.loc.gov/standards/vracore/VRA_Core4_Intro.pdf.

76. "VRA Core: A Data Standard for the Description of Images and Works of Art and Culture," Library of Congress, http://www.loc.gov/standards/vracore/.

77. "Art & Architecture Thesaurus Online," The Getty, http://www.getty.edu/research/tools/vocabularies/aat/index.html.

78. The instruction in CCO can be applied to VRA Core.

79. Patricia Harpring, "Categories for the Description of Works of Art," eds. Murtha Baca and Patricia Harpring; revised March 2014 by Patricia Harpring (Los Angeles: The Getty, 2014), http://www.getty.edu/research/publications/electronic_publications/cdwa/.

80. Patricia Harpring, "Categories for the Description of Works of Art: Overview of Categories," edited Murtha Baca and Patricia Harpring; revised March 2014 by Patricia Harpring. The Getty, http://www.getty.edu/research/publications/electronic_publications/cdwa/categories.html.

81. "CDWA Lite," The Getty, http://www.getty.edu/research/publications/electronic_publications/cdwa/cdwalite.html.

82. EDItEUR, "ONIX for Books," http://www.editeur.org/11/Books/.

83. Jean Godby, *A Crosswalk from ONIX Version 3.0 for Books to MARC 21* (Dublin, OH: OCLC Research, 2012), http://www.oclc.org/research/publications/library/2012/2012-04.pdf (report) and http://www.oclc.org/research/publications/library/2012/2012-04a.xls (crosswalk).

84. EDItEUR, "ONIX for Serials," http://www.editeur.org/17/ONIX-for-Serials/.

85. *American National Standard for Library and Information Sciences and Related Publishing Practices—Basic Criteria for Indexes, ANSI Z39.4–1984* (New York: American National Standards Institute, 1984).

86. James D. Anderson, *Guidelines for Indexes and Related Information Retrieval Devices* (Bethesda, MD: NISO Press, 1997).

87. International Organization for Standardization, *Information and Documentation: Guidelines for the Content, Organization and Presentation of Indexes*, ISO 999:1996 (Geneva: ISO, 1996).

88. International Organization for Standardization, *Information and Documentation—Guidelines for Bibliographic References and Citations to Information Resources,* ISO 690:2010, 3rd ed. (Geneva: ISO, 2010).

SUGGESTED READINGS

Baca, Murtha. "A Picture Is Worth a Thousand Words: Metadata for Art Objects and Their Visual Surrogates." In *Cataloging the Web: Metadata, AACR, and MARC 21,* edited by Wayne Jones et al., 131–38. Lanham, MD: Scarecrow Press, 2002.

Baca, Murtha, Patricia Harpring, Elisa Lanzi, Linda McRae, and Ann Whiteside. *Cataloging Cultural Objects: A Guide to Describing Cultural Works and Their Images.* Chicago: American Library Association, 2006. Also available at http://cco.vrafoundation.org/.

Chan, Lois Mai, and Athena Salaba. *Cataloging and Classification: An Introduction.* 4th ed. Lanham, MD: Rowman & Littlefield, 2016. Chapters 5–6.

Describing Archives: A Content Standard (DACS). 2nd ed. Chicago: Society of American Archivists, 2013. Also available at http://www2.archivists.org/groups/technical-subcommittee-on-describing-archives-a-content-standard-dacs/dacs.

Dublin Core Metadata Initiative. http://dublincore.org/.

"EAD: Encoded Archival Description." Library of Congress. http://www.loc.gov/ead/.

Guenther, Rebecca S. "MODS: The Metadata Object Description Schema." *Portal: Libraries and the Academy* 3, no. 1 (2003): 137–50.

Harpring, Patricia. *Categories for the Description of Works of Art*, edited by Murtha Baca and Patricia Harpring. Revised March 2014 by Patricia Harpring. Los Angeles: The Getty, 2014. http://www.getty.edu/research/publications/electronic_publications/cdwa/.

Hart, Amy. *RDA Made Simple: A Practical Guide to the New Cataloging Rules*. Santa Barbara, CA: Libraries Unlimited, 2014.

International Federation of Library Associations and Institutions. IFLA Study Group. *Functional Requirements for Bibliographic Records: Final Report.* Munich: Saur, 1998. https://www.ifla.org/publications/functional-requirements-for-bibliographic-records.

International Standard Bibliographic Description (ISBD). Consolidated ed. Berlin: De Gruyter Saur, 2011. http://www.ifla.org/files/assets/cataloguing/isbd/isbd-cons_20110321.pdf.

Intner, Sheila S., and Jean Weihs. *Standard Cataloging for School and Public Libraries.* 5th ed. Santa Barbara, CA: Libraries Unlimited, 2015. Chapter 3: "Preparing Bibliographic Descriptions."

Joudrey, Daniel N., Arlene G. Taylor, and David P. Miller. *Introduction to Cataloging and Classification*. 11th ed. Santa Barbara, CA: Libraries Unlimited, 2015. Chapter 4: "RDA Basics" and Chapter 5: "Manifestations and Items."

Maxwell, Robert L. *FRBR: A Guide for the Perplexed.* Chicago: American Library Association, 2008.

Maxwell, Robert L. *Maxwell's Handbook for RDA, Resource Description & Access: Explaining and Illustrating RDA: Resource Description and Access Using MARC21*. Chicago: American Library Association, 2013.

Moulaison, Heather Lea, and Raegan Wiechert. *Crash Course in Basic Cataloging with RDA*. Santa Barbara, CA: Libraries Unlimited, 2015. Chapter 6: "RDA: Manifestations and Items."

Svenonius, Elaine. *The Intellectual Foundation of Information Organization.* Cambridge, MA: The MIT Press, 2000. Chapter 7: "Document Languages."

Taylor, Arlene G. "An Introduction to Functional Requirements for Bibliographic Records (FRBR)." In *Understanding FRBR: What It Is and How It Will Affect Our Retrieval Tools,* edited by Arlene G. Taylor, 1–19. Westport, CT: Libraries Unlimited, 2007.

Text Encoding Initiative. "TEI: P5 Guidelines." Version 3.1.0. Last updated on December 15, 2016. http://www.tei-c.org/release/doc/tei-p5-doc/en/html/index.html.

"VRA Core: A Data Standard for the Description of Images and Works of Art and Culture." Library of Congress. https://www.loc.gov/standards/vracore/.

CHAPTER 8

ACCESS AND AUTHORITY CONTROL

The issues addressed in this chapter, related to access and authority control, are expressed in the following questions: How are resources made available to users? How can identification of relationships among entities assist users in finding what they need? How do access points and authority control affect collocation? How can entities with the same name be distinguished from each other? How can variant names that represent the same entity be brought together? How can all versions of the same work be brought together? What standards and tools are available to assist with the processes of providing access and authority control?

ACCESS

In most online systems keyword searching is available. If a user knows exactly what words or names have been used in a description, then keyword searching is successful. But users who do not know exact words or names have to guess. This is sometimes successful and sometimes not. Why are keywords a problem in search systems? For a variety of reasons:

- Most words in the English language have more than one meaning, and most concepts have more than one word that represents them (discussed further in Chapter 10).
- Different users do not think of the same word(s) to express a concept, and different creators do not necessarily use the same word(s) to write about a concept.
- Persons and families do not necessarily retain the same name or same form of name throughout their existences, and corporate bodies do not necessarily use the same name or form of name in their documents, nor are they known by the same form of name by everyone.
- Different people, families, and corporate bodies have the same name, and many different corporate bodies have the same acronym.
- Titles of works that are reproduced, edited, or adapted are not always the same in the original and the new version.
- Titles are not unique; the same title can be given to very different works.

For all of these reasons and more, libraries, archives, and museums came to the realization that there needs to be consistency and predictability in retrieval systems, something that goes beyond the luck of keyword searching. An alternative to relying on keywords is to have specific formal *access points* (i.e., names, titles, subjects, etc., chosen by a cataloger or indexer and placed in particular fields in a record) provided in metadata descriptions. These access points need to be expressed consistently from record to record when several different records use the same word or name as an access point. In some systems, one is often designated as the main or primary access point.

The rest of the metadata world has also begun to see a need for access points either with consistent form from record to record or with links to resources that uniquely identify a person, corporate body, concept, and so forth. Much energy in the last three decades was spent on determining what descriptive metadata is needed for particular forms of information resources and on how these descriptions can be encoded for display. As metadata accumulates to the massive numbers that library catalogs have been dealing with, the need for attention to problems of access to all this metadata becomes more apparent.

BIBLIOGRAPHIC RELATIONSHIPS

Few information resources exist in total isolation. The majority have various kinds of relationships to other resources. A resource may have a creator who has also written other resources. It may have the same publisher or be in the same series as other related resources. It may originate in an institution where a research project has produced multiple reports concerning the same data. It may be a manifestation of the same work but in a different format or medium. It may have intellectual content that is the same or similar to that of another resource, or it may be an adaptation, a translation, a commentary, or a teacher's guide to another work. It may be a supplement, a sequel, a part of a larger work, or conversely, it may be an entity that contains a number of smaller works within it. It may be a performance of a work that also appears in written form. Relationships abound in the bibliographic universe.

Information systems operate on the perception that a user interested in one resource may very well be interested in others that are related to it. Therefore, successful document retrieval relies on having such relationships identified, at least implicitly, and retrieval is improved when such relationships are made explicit. When relationships are merely described (e.g., mentioned in a note)—or not even described—a user might discover those relationships only by chance. Access points can make relationships explicit. Library cataloging has provided rules for such access points at least since Seymour Lubetzky identified principles upon which the Paris Principles (1961) were based. However, research to formally identify and categorize relationships began only in the 1980s. Barbara Tillett researched and presented a taxonomy of bibliographic relationships:[1]

- **Equivalence relationships**: found in exact reproductions of the same manifestation of a work; include copies, issues, facsimiles, reprints, photocopies, microforms, audiotapes of sound recordings on disc (if it is the same content), and other such reproductions
- **Derivative relationships**: found in modifications based on particular expressions or works; include new editions, revisions, adaptations, changes of genre (e.g., dramatization of a novel), new works based on style or thematic content of other works, and the like
- **Descriptive relationships**: found in descriptions, criticisms, evaluations, or reviews of a work; include book reviews, annotated editions, critiques, commentaries, and so forth
- **Whole–Part relationships**: found in a component part of a larger work or in the relationship between a work and each of its various parts; include selections from anthologies or collections, articles from journals, maps in atlases, series that have independent works within them, and so on
- **Accompanying relationships**: found in bibliographic resources that are created for the purpose of complementing particular works; they can complement equally, or one work can be the principal or predominant work; include resources with supplements (e.g., teacher's manual for a textbook), software manuals or help programs, concordances, indexes, parts of a kit, and other such complementary relationships
- **Sequential relationships**: found in bibliographic resources that continue or precede other resources; include successive titles of a serial, sequels and prequels of a movie, parts in a numbered series, and the like
- **Shared characteristic relationships**: found in any resources that coincidentally share characteristics in common, such as common creators, titles, dates, subjects, language, country of publication, and so on

Following upon Tillett's research, others have used empirical research to refine definitions and further delineate ways in which such relationships could be explicitly linked in retrieval tools.[2] For example, Richard Smiraglia presented the following taxonomy of Tillett's derivative relationships:[3]

- **Simultaneous derivations**: works published in two editions nearly simultaneously, such as a British edition and a North American edition of the same work; may have different titles, different sizes, or other differing characteristics
- **Successive derivations**: works revised one or more times and issued anew with statements indicating revision, as well as resources issued successively that may not have revision statements
- **Translations**: works presented in languages other than the original

- **Amplifications**: original works that have added to them such things as illustrations, musical settings, commentaries, and so forth
- **Extractions**: works presented in smaller forms such as abridgments, condensations, and excerpts
- **Adaptations**: resources that modify original works; for example, simplifications, screenplays, librettos, arrangements of musical works, and so on
- **Performances**: sound or visual recordings of works, each of which may differ in such things as tone, amplification, and interpretation

Making such relationships explicit requires formalized rules for creating and applying a unique identifier for a work (*a work ID*) and, possibly, connecting devices such as cross-references.[4] This, of course, requires a consistent definition of *work,* which eluded organizers and researchers until the end of the twentieth century.[5] The *Anglo-American Cataloguing Rules* (AACR) has provided such rules for many years, although they were somewhat flawed and inconsistent because of the problems with definitions. RDA, now, provides rules for naming a work. RDA's work ID, in most cases, is a standardized name/title access point, but in cases where a creator's name is unknown or not easily determined, the work ID is title only. Work IDs often require the addition of qualifiers or relationship designators in order to make relationships explicit, and references are required to bring together variant forms of names of creators and titles.

AUTHORITY CONTROL

Authority control has three major concerns: consistency, relationships, and uniqueness. It is the result of the processes of:

(1) maintaining consistency in the verbal forms used to represent entities as access points,

(2) showing the relationships among entities (particularly agents, works, and subjects), and

(3) ensuring that entities are identified uniquely and the access points chosen to represent them cannot be used for any other entity.

In other words, authority-controlled access points are consistently presented, they are explicitly linked to other related access points, and they are unique because similarly named entities have been disambiguated.

Authority control is also the result of the process of doing authority work with or without the necessity of choosing one name or title or one subject term to be the "authorized" selection. If every variant name, title, or term is given equal status, then one is chosen for default display. Whether or not an authorized form is chosen to represent the entity, a searcher may use any of the forms in the authority record to gain access to information resources related to the entity. The process called *authority control* was so named because it was thought necessary to determine an authorized form for every entity known by variant forms. For example, the English form of Confucius's name has been determined by the Library of Congress (LC) to be the *preferred name* (formerly known as *authorized form of name*) in the United States, and the names given to this person in many other languages, including his Chinese name for himself, have been relegated to positions of variant forms that act as references to the authorized form.

The preferred name is the basis for an *authorized access point* (AAP); an AAP is a human-readable, established string of alphanumeric characters used to represent an entity. It is generally documented in an authority record, which is entered into a retrieval tool to help ensure consistency in retrieval. *Variant names* are the basis for *variant access points* (VAPs), which are found in authority records as well. VAPs are also strings that represent an entity, but they are *not* authorized for use in bibliographic metadata. The following is an example of some of the data found in an authority record.

AAP:	**Shakespeare, William, 1564–1616**
VAPs:	Shakspeare, William, 1564–1616 Szekspir, Wiliam, 1564–1616 Saixpēr, Gouilliam, 1564–1616 Ṣēkspiyar, Viliyam, 1564–1616 Tsikinya-chaka, 1564–1616

Шекспир, Вильям, 1564–1616
沙士比亞威廉姆, 1564–1616
[many others]

If this data is included in a retrieval tool, it means that any search on an alternate form of name will cause a reference to be displayed to the searcher. For example, a search for *Szekspir, Wiliam* could retrieve the following message:

"Szekspir, Wiliam, 1564–1616" is not used in this library's catalog. **Shakespeare, William, 1564–1616** is used instead.

Try a search for **Shakespeare, William, 1564–1616.**

As the world of information becomes more global, many have recognized that different forms of a name (e.g., in different languages or scripts) can be used for access. In addition, the concepts of *authority* and *control* in a culture with an emphasis on individualism are not readily welcomed, and the words have negative connotations for many people. Efforts were made by Arlene G. Taylor and others a few years ago to begin using the phrase *access control* instead of *authority control.* But *access control* has come to mean a computer operating system feature that controls the access that certain categories of users have to files and to functions. (It has also come into use in some airports to name the function of determining who may enter a country.) We have not yet found better terminology.

Goals for authority control are to assist users to

- find an entity (e.g., creator, place, subject, work) using the searcher's vocabulary; for example, to find books by Hans Christian Andersen, even if he used H.C. Andersen on the title page; to find resources about *bathing suits* using the search term *swimsuits* or *bikinis*;
- identify entities; for example, to identify the Michael Gorman who is a librarian, as opposed to the architect, the fiddle player, or the fly-fishing enthusiast;
- collocate resources, persons, and other entities regardless of terminology used; for example, the retrieval tool brings together all resources related to the opera *La Traviata*, including Verdi's score, a separately published libretto by Piave, various sound recordings, and so on;
- aid in evaluating or selecting the appropriate resource or the entity identified (e.g., Is Vincent Willem van Gogh the same person as the artist Vincent van Gogh? Is the use of *bridge* in a search limited to the card game, or will resources about structures across rivers or about dental work be retrieved?);
- explore relationships among names and entities (e.g., How are Mark Twain, Samuel Langhorne Clemens, Quintus Curtius Snodgrass, and Louis de Conte related?); and
- provide synonyms and syndetic structure to aid in subject searching or to provide variant forms of name for persons, families, corporate bodies, or works.

Authority control is needed for collocation—for bringing together everything related to a person, family, corporate body, place, work, or expression, regardless of what name has been used for one of those entities. One has only to look at some websites that do not have authority control to see that there is value in collocation of forms of name. For example, a recent search for *Smith, Susan* in the books category at Amazon.com[6] resulted in the following first 12 out of 2,257 results sorted by Amazon's default arrangement of results (relevance):

1. *The History of Joseph* ***Smith*** *by His Mother* by Lucy Mack **Smith** and **Susan** Easton Black
2. *Time Expired* (Jill **Smith** Mysteries, Book 8) by **Susan** Dunlap
3.–7. [Five more titles by **Susan** Dunlap in the Jill **Smith** Mystery series]
8. *Caring Liturgies: The Pastoral Power of Christian Ritual* by **Susan** Marie **Smith**
9. *A Vampire's Promise* by Carla **Susan Smith**
10. *Death and Taxes* (Jill **Smith** Mysteries, Book 7) by **Susan** Dunlap
11. *The Book of Jeremiah: The Life and Ministry of Jeremiah A. Wright, Jr.* by **Susan** Williams **Smith**
12. *My Daughter* ***Susan Smith*** by Linda Russell and Shirley Stephens

One can see that the first seven results are not resources created by someone named *Susan Smith*. There is a mixture of different types of keywords that were used to retrieve these results (e.g., parts of names, titles proper, series titles). When an author search was performed in Amazon's advanced search screen, the results were more focused, but still lacked consistency and useful collocation:

1. *A Vampire's Hunger* by Carla **Susan Smith**
2. *Early Childhood Mathematics* (5th ed.) by **Susan** Speery **Smith**
3. *The Book of Jeremiah: The Life and Ministry of Jeremiah A. Wright, Jr.* by **Susan** Williams **Smith**
4. *A Vampire's Soul* by Carla **Susan Smith**
5. *How to Become a Million Dollar Real Estate Agent in Your First Year: What Smart Agents Need to Know Explained Simply* by **Susan Smith** Alvis
6. *A Vampire's Honor* by Carla **Susan Smith**
7.–1,855. [Many other resources created by or contributed to by someone with a name containing *Susan* and *Smith*]

On the other hand, a search for *Smith, Susan* in the LC catalog[7] on the same day yielded 86 much more focused results:

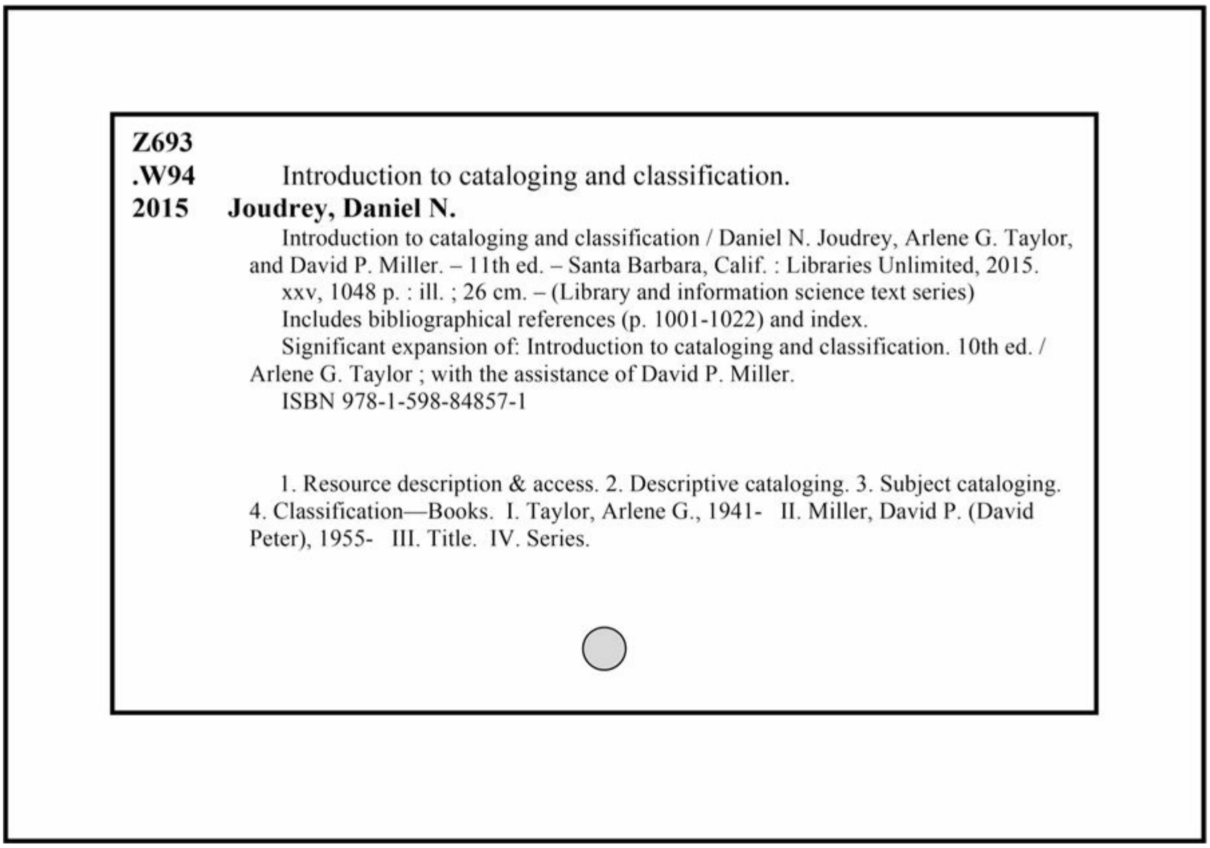
Smith, Susan (20 [records associated with this name])
Smith, Susan, 1885– (8)
Smith, Susan, 1937– (0)
    See: **Smith, S. A. M., 1937–**
Smith, Susan, 1940– (0)
    See: **Smith, Susan E. (Susan Elizabeth), 1940–**
Smith, Susan, 1941– (1)
Smith, Susan, 1944– (1)
    . . .
Smith, Susan, 1963– (2)
Smith, Susan A. (2)
Smith, Susan Arnout (3)
    . . .
Smith, Susan Vernon (4)
    See Also [the pseudonym]: **Enfield, Carrie**
    See Also [the pseudonym]: **Vernon, Rosemary**
Smith, Susan Way- (0)
    See: **Way-Smith, Susan**
Smith, Susan Welsh- (0)
    See: **Welsh-Smith, Susan**
Smith, Susan Williams, 1954– (0)
    See: **Smith, Susan K. Williams, 1954–**
Smith, Susan Williamson (1)

The preceding results were retrieved by using the browse function for creators; other search functionality naturally will provide different kinds of results. One may not know from this list, at first, which Susan Smith is which, but once a needed Susan Smith has been identified, then all titles relating to that person are found together, and it is not necessary to look at any of the other entries.

Authority Work

In order to have authority control it is necessary for someone to do authority work. This requires identifying all variants for a name or title of a work and making the necessary decisions about which variants represent the same names, which should be the authorized form, and which should be references. Most systems still require the choice of one name as the *authorized access point* (formerly known as the *heading*). The entries from the LC catalog above show AAPs in boldface type, VAPs in plain type, with *see* and *see also* labels in italics. Current practice dictates the establishment of an AAP for each name or work title that is intended to be an access point. The term *heading* came from print catalog days when each access point was printed at the top (head) of the copy of the description and was called the heading for the record (see Figure 8.1). In the online catalog, the AAP no longer appears at the head of each record, although sometimes it is at the head or the side of a list of records in a catalog display (see Figure 8.2).

The process of authority work involves more than determining the authorized name or title, however. With the recent move to *RDA: Resource Description & Access*, the process also regularly includes recording significant descriptive information that assists in identifying the entity. For example, in addition to sorting out a person's use of names, those performing authority control include metadata about places of birth and death, affiliations, field(s) of activity, and so on (see Figure 8.3).



Z693
.W94 Introduction to cataloging and classification.
2015 **Joudrey, Daniel N.**

Introduction to cataloging and classification / Daniel N. Joudrey, Arlene G. Taylor, and David P. Miller. – 11th ed. – Santa Barbara, Calif. : Libraries Unlimited, 2015.
xxv, 1048 p. : ill. ; 26 cm. – (Library and information science text series)
Includes bibliographical references (p. 1001-1022) and index.
Significant expansion of: Introduction to cataloging and classification. 10th ed. / Arlene G. Taylor ; with the assistance of David P. Miller.
ISBN 978-1-598-84857-1

1. Resource description & access. 2. Descriptive cataloging. 3. Subject cataloging. 4. Classification—Books. I. Taylor, Arlene G., 1941- II. Miller, David P. (David Peter), 1955- III. Title. IV. Series.

Figure 8.1. Catalog Card Showing the Title Placed at the Top of the Record as the Heading for the Card.

Number	Author / Titles	Media	Year
1	**Smith, Susan**		
	Building A Successful Career In Counseling And Psychotherapy.	Streaming Video	2016
	Sweating Indian Style: Conflicts Over Native American Ritual	Streaming Video	2014
2	**Smith, Susan, 1885-**		
	The Christmas Tree In The Woods	Book	1932
	Made in America	Book	1929
	Made in France	Book	1931
3	**Smith, Susan, 1946-**		
	Communications In Nursing : Communicating Assertively & Responsibly In Nursing	Book	1996
4	**Smith, Susan, 1963-**		
	The Musical : Race, Gender And Performance	Book	2005
5	**Smith, Susan Belasco, 1950-**		
	Periodical Literature In Nineteenth-Century America	Book	1995
6	**Smith, Susan E, 1953-**		
	Fear Or Freedom : A Woman's Options In Social Survival And Physical Defense	Book	1986
7	**Smith, Susan Harris**		
	American Drama : The Bastard Art	Book	1997
	Plays In American Periodicals, 1890-1918	Book	2007
8	**Smith, Susan Jane, 1945-**		
	Organizing A Local Government Documents Collection	Book	1979
9	**Smith, Susan L. (Susan Louise), 1947-**		
	The Power Of Women : A Topos In Medieval Art And Literature	Book	1995
10	**Smith, Susan Sharpless**		
	Web-Based Instruction : A Guide For Libraries	Book	2010

Figure 8.2. A Representation of Search Results Showing the Authorized Access Points for Various Persons Named *Susan Smith*.

ARN 348864

Rec stat	c	Entered	19800602	Replaced	20121016074856.0		
Type	z	Upd status	a	Enc lvl	n	Source	
Roman	|	Ref status	b	Mod rec		Name use	a
Govt agn	|	Auth status	a	Subj	a	Subj use	a
Series	n	Auth/ref	a	Geo subd	n	Ser use	b
Ser num	n	Name	a	Subdiv tp	n	Rules	z

010 n 79117152
040 DLC $b eng $e rda $c DLC $d DLC $d MWA $d DLC $d OCoLC ...
046 $f 18321129 $g 18880306
053 _0 PS1015 $b PS1018
100 1_ Alcott, Louisa May, $d 1832-1888
370 Philadelphia, Pa. $b Boston, Mass.
374 Author
375 female
377 eng
400 1_ Ol'kot, Luiza, $d 1832-1888
400 1_ Alkūt, Luwīzā, $d 1832-1888
400 1_ Barnard, A. M., $d 1832-1888
400 1_ Alcott, Louisa M. $q (Louisa May), $d 1832-1888
400 1_ Alcott, L. M. $q (Louisa May), $d 1832-1888
400 0_ Author of Little women, $d 1832-1888
400 1_ Little women, Author of, $d 1832-1888
400 0_ Author of An old-fashioned girl, $d 1832-1888
400 1_ Old-fashioned girl, Author of, $d 1832-1888
400 0_ Author of Little men, $d 1832-1888
400 1_ Little men, Author of, $d 1832-1888
...
400 0_ Олкотт, Луиза Мэй, $d 1832-1888
400 1_ アルコツトルイザメイ, $d 1832-1888
400 1_ オルコツトルイザメイ, $d 1832-1888
400 1_ ルイザメイオルコット, $d 1832-1888
667 Machine-derived non-Latin script reference project.
667 Non-Latin script references not evaluated.
670 Gulliver, L. Louisa May Alcott, a bibliography ... 1932.
670 nuc88-133929: Her Hospital sketches [MI] 1863 $b (hdg. on NNC-T rept.: Alcott, Louisa May, 1832-1888; usage: L.M. Alcott)
670 Her Jo's boys, 1983: $b t.p. (Louisa M. Alcott)
670 MWA/NAIP files $b (hdg.: Alcott, Louisa May, 1832-1888; usage: Louisa M. Alcott; author of Little women; author of An old-fashioned girl; author of Little men; author of Hospital sketches; author of Kitty's class-day; author of Moods; author of Aunt Jo's scrap-bag; author of Work; author of Aunt Kipp; author of Eight cousins; author of Psyche's art)
670 Wikipedia, Dec. 9, 2010 $b (Louisa May Alcott; b. Nov. 29, 1832 in Germantown (Philadelphia), Pa.; d. Mar. 6, 1888 in Boston, Mass.; American novelist)
678 0_ Louisa May Alcott (1832-1888) was an American novelist.

Figure 8.3. An Excerpt of an Authority Record. (Source: OCLC Connexion, Authorities—Record Number 348864)

Authority work also includes identification of relationships of all variant names and titles to their corresponding authorized forms, or between two or more related names or titles (e.g., the relationship between the pseudonym *Mark Twain* and the birth name *Samuel Langhorne Clemens*). For example, in a catalog, authority data triggers the following *see also* reference based on a search for **Twain, Mark**:

For works of this author written under other names, search also under:
Clemens, Samuel Langhorne, 1835–1910;
Snodgrass, Quintus Curtius, 1835–1910;
Conte, Louis de, 1835–1910; and
Alden, Jean François, 1835–1910.

Once decisions about the entity have been made, authority work then involves the process of documenting, in an authority record, the work done along with the decisions made. An authority record is a compilation of metadata about a person, a family, a corporate body, a place, a work, an expression, or a subject. It includes evidence of all the decisions made and all the relationships among variants that have been identified.

Thinking in terms of global or international access allows the possibility for authority records to contain all variant forms for an entity, without designating one as the authorized one. In any case, a search for any one of the variants for an entity should retrieve all metadata records associated with the entity. In the process of authority work, even though there may be a desire to provide more global access by not designating one of the variants as the authorized one, it is still necessary to go through the process of creating an authorized access point in particular locales so that one form can be designated for default display for cases where a user does not wish to designate a preference. Doing authority work in concert with many other people is very helpful, saving hours of work for individual catalogers. One such cooperative project is NACO, the Name Authority Cooperative Program of the PCC (Program for Cooperative Cataloging).[8] Participants in NACO follow a common set of standards and guidelines when creating or updating authority records so that consistency is maintained in the Library of Congress/NACO Authority File (LCNAF; formerly known as the Library of Congress Name Authority File), a large shared authority file.

Authority Files

In earlier chapters we said that metadata models are built on the proposition that resources are entities that are connected to other entities (e.g., creators, contributors, subjects). Entities have attributes, and those attributes have values, some of which are controlled (refer back to Figure 5.6 on page 228). With authority control we maintain consistency over the values provided in *data value standards*. One traditional name for a data value standard is *authority file.* Authority records are accumulated and stored in an authority file or files, which are separate from the bibliographic files (which contain metadata for individual resources). In most modern integrated library systems, bibliographic records are often linked to the authority records that contain metadata for names, works, and subjects (refer back to Figure 4.1 on page 170 for an example of linkages). If linked, references from authority records can be displayed to users in the results of their searches of the bibliographic file, such as:

Smith, Susan, 1940–
See: **Smith, Susan E. (Susan Elizabeth), 1940–**

In the United States, one of the most widely used authority files is the LCNAF. Other data value standards in use include subject heading lists, thesauri, and classification schemes, which are discussed in later chapters. Several different data value standards or authority files exist for the purpose of name control. Besides the LCNAF these include, among others, the Getty vocabularies, which are used primarily in the art and museum community. In recent years, many data value standards have been made available in linked-data compatible formats. This entails providing the same authority data but with *uniform resource identifiers* (URIs) or *internationalized resource identifiers* (IRIs) that provide unambiguous identification. For example, the following are just a few of the authority files and controlled vocabularies in LC's Linked Data Service:[9]

- *American Folklore Society Ethnographic Thesaurus*
- Extended Date/Time Format
- LCNAF
- *Library of Congress Classification* (LCC)
- *Library of Congress Demographic Group Terms* (LCDGT)

- *Library of Congress Genre/Form Terms for Library and Archival Materials* (LCGFT)
- *Library of Congress Subject Headings* (LCSH)
- MARC Relator, Country, Language, and Geographic Area codes
- RDA Content, Media, and Carrier types

The data is provided in a variety of formats (e.g., RDF in XML, N-Triples, and JSON), using several different data structure schemas (e.g., Metadata Authority Description Schema [MADS], Simple Knowledge Organization System [SKOS], and MARCXML) in order to be available for wide use. This freely available service could allow for authority control concepts to be applied beyond the library and information science (LIS) communities of practice—even into the general web environment—but this will happen only if online content creators take advantage of this authoritative and trustworthy data.

International Authority Control Efforts

Work is proceeding toward international authority control. A project that has made some headway in assisting in this goal is the *Virtual International Authority File* (VIAF). Authority work is done all over the world. National bibliographic agencies have independent authority files, each of which contains data about the same entities that also appear in other files. This highly trusted data that has been created over many years by national libraries and bibliographic agencies, however, is distributed widely. At the turn of the century, the VIAF project began to bring this data together in one interface.

> VIAF's goal is to make library authority files less expensive to maintain and more generally useful to the library domain and beyond. To achieve this, VIAF seeks to include authoritative names from many libraries into a global service that is available via the Web. By linking disparate names for the same person or organization, VIAF provides a convenient means for a wider community of libraries and other agencies to repurpose bibliographic data produced by libraries serving different language communities.[10]

In the design phase of VIAF, there were several suggestions as to how to bring this data together. The method chosen was a distributed model, based on a suggestion by Barbara Tillett, instead of creating one massive authority file.[11] The Library of Congress, OCLC, and the German national library established VIAF as a joint project; OCLC maintains the service and provides open access to the data for everyone.[12] At the time of this writing, VIAF contains authority data from 37 agencies in 29 countries for persons, corporate bodies, families, works, expressions, meetings, and geographic names. It displays national and regional variations in authority data based on differences in language, character sets, and spelling.[13]

We do not yet know exactly where initiatives like VIAF will lead. One possibility is that all forms of name could be accumulated in one authority display such as VIAF, with none chosen as *the* one that must be used. Each form would be identified by language, the national library responsible, or rule set. A search for any one of the forms would retrieve records related to the entity for which information is being sought. A difficulty with this scenario is the immense size of records that would be produced for many names. Authority records already can have multiple forms of name even in one language; the addition of multiple forms in each of multiple languages could result in *very* large records (much larger than what is seen in Figure 8.3).

International authority control requires standards, and there are several that may be used to good advantage. In the international library community, the International Federation of Library Associations and Institutions (IFLA) issued its second edition of *Guidelines for Authority Records and References (GARR)* in 2001;[14] plans to update it again are in the works as of this writing. *Functional Requirements for Authority Data* (FRAD)[15] and *Functional Requirements for Subject Authority Data* (FRSAD)[16] also address authority control issues; both, however, are being superseded by the *IFLA Library Reference Model* (LRM).[17] The Z39.50 protocol (described in Chapter 4), as well as the movement to place more bibliographic and authority data online, makes it possible to retrieve and display authority data quickly and efficiently. This enables the sharing of authority data for the same entities in different languages, and forms of name in any particular language can be verified by experts in that language. The OAI (Open Archive Initiative)[18] promotes a protocol (Open Archives Initiative Protocol for Metadata Harvesting, or OAI-PMH) that defines a mechanism for harvesting XML-formatted metadata. It is intentionally simple with the intent of being easier and less costly to implement than previous communications protocols (e.g., Z39.50). It can be used for authority, as well as bibliographic, metadata.

GENERAL BIBLIOGRAPHIC STANDARDS FOR ACCESS AND AUTHORITY CONTROL

As mentioned in Chapter 7, the creation of metadata in a broad sense requires description of the resource, attention to access points for the description, and encoding. Encoding is discussed in Chapter 6, and Chapter 7 addresses description. The next step after description is the selection of access points that enable consistent and predictable retrieval. In order to have names, titles, and subjects both collocated and differentiated, the access points need to be brought under authority control.

A number of models and standards address the issues of access and authority control in libraries to varying degrees. First we discuss some general bibliographic models and standards; standards that exist for archives and for the art and museum communities are described in later sections. The library-oriented standards include the conceptual models associated with IFLA's *Functional Requirements* documents (FRBR, FRAD, FRSAD, and LRM), as well as

- *Statement of International Cataloguing Principles* (ICP)
- *RDA: Resource Description & Access*
- *Anglo-American Cataloguing Rules, Second Edition* (AACR2)
- Metadata Authority Description Schema (MADS)

IFLA's "Functional Requirements" Family of Standards

Functional Requirements for Bibliographic Records (FRBR), discussed in Chapter 5, was IFLA's original attempt to formalize the entities, attributes, and relationships that comprise the bibliographic universe. FRBR defines three groups of entities. Group 1 is focused on information resources—products of creation or intellectual endeavor. Resources comprise four entities or levels: *work*, *expression*, *manifestation*, and *item* (WEMI). Attributes for each entity (e.g., title of manifestation, language of expression) may be included in the descriptive metadata for a resource. Two of these Group 1 entities, *work* and *expression*, also may be described in authority data, with AAPs established as unambiguous identifiers for them. For example, the work Shakespeare's *Hamlet* has been known by different titles across various editions and translations. It is, therefore, helpful to have a consistent way to identify the work (or a particular expression of that work) using an AAP such as **Shakespeare, William, 1564–1616. Hamlet**.

In addition, FRBR identifies Group 2 entities—agents responsible for creating the content of resources or otherwise contributing to them. This group includes *person* and *corporate body,* with *family* added later by FRBR's companion model, FRAD, in 2009.[19] A third group in the original FRBR model identifies entities that can be subjects of works—this third group comprises *concept, object, event,* and *place,* plus the entities from Groups 1 and 2. An investigation, which focused on more clearly delineating problems associated with subject entities, was published as FRSAD in 2010.[20] This third model did *not* address individual entities for specific kinds of subjects (e.g., *concept*, *place*, *person*), focusing instead on a more generic level (see discussion below).

It is fairly clear that the Group 2 and Group 3 entities are meant to be access points, but the FRBR model does not describe how those entities and relationships are reflected in access points in bibliographic records, nor does it delineate how a cataloger is to determine the name or AAP that is to be used for these entities.[21] The task of addressing authority-controlled data was assigned to the IFLA Working Group on Functional Requirements and Numbering of Authority Records (FRANAR), which devised FRAD, and the Working Group on Functional Requirements for Subject Authority Records (FRSAR), which designed the FRSAD model.

Functional Requirements for Authority Data (FRAD)

FRAD, like FRBR, is a conceptual model. Its purpose is to provide a framework for the authority data necessary to support authority control both locally and internationally. It was never meant to be a practical application or implementation plan. FRAD, like FRBR, is centered on users, aiming to place in authority records the data that is required by the users of those records. Users of authority records are defined to be both the authority record creators and the end users who benefit both from authority information being available to them as reference tools and from improved retrieval tools that contain controlled access points with reference structures. Four user tasks are identified: *find, identify, contextualize,* and *justify.* The tasks *find* and *identify* are the same as in FRBR. The tasks *contextualize* and *justify* are clearly meant to apply to authority record creators, not to end users. The term *contextualize* means to place in context or to clarify relationships; *justify* means to document one's reason for choosing the form of name on which a controlled access point is based.

The FRAD model, as mentioned already, added *family* to FRBR's Group 2 entities. In addition, FRAD added five entities needed for a complete model of authority control: *name, identifier, controlled access point, rules,* and *agency.* These may be defined as follows:

- **Name**: the word, character, group of words, or character string that an entity is known by (e.g., a personal name, a title).
- **Identifier**: anything that uniquely identifies an entity, such as a number, a code, a word or phrase, a logo, or other such device (e.g., no2006073134).
- **Controlled access point**: a name, term, code, and the like, that is designated in a bibliographic record, authority record, or reference record as a character string to be used for retrieval (e.g., **Dickinson, Emily, 1830–1886**).
- **Rules**: a set of instructions for formulating controlled access points (e.g., RDA).
- **Agency**: an organization that has responsibility for applying rules to the construction and/or modification of controlled access points (e.g., LC or another library).

FRBR identifies some attributes for each of the entities in its model. FRAD adds many attributes of its own. For example, the attributes enumerated in FRAD for *person* are listed below; the ones with an asterisk are unique to FRAD.

- Dates associated with the person
- Title of person
- Other designation associated with the person
- Gender*
- Place of birth*
- Place of death*
- Country*
- Place of residence*
- Affiliation*
- Address*
- Language of person*
- Field of activity*
- Profession/Occupation*
- Biography/History*

FRAD, like FRBR, uses entity-relationship modeling techniques to show the relationships between and among the entities of the model. Both models also identify the major types of relationships that operate between instances of the same entity type and between instances of different entity types. FRAD has an extensive relationships section, providing discussion of (1) relationships that operate at a generic level between the entity types, (2) relationships that are commonly shown in the reference structure of the authority record, and (3) relationships between one controlled access point and another.[22] For example, the following table illustrates relationships between entity types:[23]

Entity Type	**Sample Relationship Types**

Person to Person	• **Pseudonymous relationship** • **Attributive relationship** • **Collaborative relationship** • **Sibling relationship** • **Parent/child relationship**
Person to Family	• **Membership relationship**
Person to Corporate Body	• **Membership relationship**
Family to Family	• **Genealogical relationship**
Corporate Body to Corporate Body	• **Hierarchical relationship** • **Sequential relationship**
Work to Work	• **Equivalence relationship** • **Derivative relationship** • **Descriptive relationship** • **Whole–Part relationship** • **Accompanying relationship** • **Sequential relationship** • **Shared characteristic relationship**

Part II of FRAD delineates, in some detail, the practice of providing authority data—mostly within the library sector, although brief mention is made of the archives, art, cultural heritage, and museum communities.

IFLA Library Reference Model (LRM)

It can be difficult to keep two or more related models in sync when they have been developed and published over an extended period of time and were completed by different committees with different charges. This was the case with FRBR, FRAD, and FRSAD. Although they are part of the same family of "Functional Requirements" models, they have different approaches to some concepts. For example, they each approach the concepts of *names* and *subjects* differently. FRBR identifies *name* as an attribute of an entity, and for *subjects*, it provides Group 3 entities (*concept*, *object*, etc.) that have relationships to the Group 1 entity *work*. FRAD treats *name* as a separate entity (e.g., a person has a relationship to a name) but treats *subject* as an attribute of *work*. FRSAD, which addresses subjects only, approaches the concept at a more abstract level. It introduces two new terms for familiar entities: *nomen*, meaning a name or label, and *thema*, meaning subject matter. For conceptual consistency within the cataloging world, these types of discrepancies had to be addressed. This is the purpose of developing LRM: to address inconsistencies and harmonize the models.

In the draft version of LRM, changes were introduced. LRM is still an entity-relationship model of the bibliographic universe, but it is narrower in scope. Distinctions between bibliographic and authority data are de-emphasized, as articulated in LRM's scope statement:

> The IFLA LRM model aims to make explicit general principles governing the logical structure of bibliographic information, without making presuppositions about how that data might be stored in any particular system or application. As a result, the model does not make a distinction between data traditionally stored in bibliographic or holdings records and data traditionally stored in name or subject authority records. For the purposes of the model, all of this data is included under the term bibliographic information and as such is within the scope of the model.[24]

Some basic entities have changed in the harmonized model. FRBR and FRAD's specific Group 2 entities are simply referred to as *agents* in LRM. Agents play roles in creating, distributing, manufacturing, owning, and modifying resources. Agents, therefore, are logical access points to include in bibliographic descriptions. An agent may be a *person* or a *collective agent* (an entity that subsumes *family* and *corporate body*). Another change is that the definition of *person* no longer allows for fictitious, legendary, and mythological characters, as well as individual nonhuman entities (e.g., Lassie, the dog; Keiko, the whale) to be viewed as persons (which was possible under FRBR). LRM also changes how subjects are addressed. It has eliminated the four unique subject entities from FRBR—*concept*, *object*, *event*, and *place*—in favor of the more generic approach found in FRSAD, but with some additional changes. *Thema* has been replaced in LRM with an even more generalized entity, *res* (i.e., *thing* in Latin); everything in the universe is a *res*. When it comes to subject matter, the two key relationships explicitly identified in the model are that (1) a *res* can be the subject of a *work* and (2) a *res* has an appellation, which is a *nomen*.

The LRM revision process has not been kind to FRAD. Most of the FRAD entities have been deprecated. *Rules* and *agency* are considered out of scope because of their administrative focus. FRAD's *name* entity has been merged into FRSAD's *nomen*. The entities *identifier* and *controlled access point* are also subsumed under *nomen*. The FRAD-specific user tasks (*contextualize* and *justify*) have been removed entirely from the LRM. Both tasks are administrative in nature and are now considered "out of scope" for the conceptual framework.[25]

Many of the FRAD relationships identified in the previous section have been generalized in LRM. For example, a pseudonymous person-to-person relationship is represented by a res/nomen relationship (*res* has appellation *nomen*) and a descriptive work-to-work relationship is now represented by a work/res relationship (*work* has subject *res*). The list of relationships specified in LRM is found in Table 5.6 on page 220. More details on specific relationships and the rest of the changes in the harmonized model can be found in the LRM documentation.

RDA has been greatly influenced by the original FRBR and FRAD models (FRBR more so than FRAD). It is clear, however, that the changes brought forth in the LRM are expected to influence the future directions of RDA, especially its overall structure. The changes in user tasks, entities, attributes, and relationships will eventually find their way into the current library content standard (changes are expected to appear in 2018). *Exactly* how these changes will be manifested, though, is unknown at the time of this writing.

Statement of International Cataloguing Principles

The *Statement of International Cataloguing Principles* (ICP),[26] published in 2009 and updated in 2016, is IFLA's revision of its 1961 Paris Principles. The ICP provides guidelines that are applicable to the online library catalogs of today as well as to the retrieval tools of the future. The revision process consisted of a series of IFLA Meetings of Experts on an International Cataloguing Code, the fifth and last of which was held in 2007.[27] These new principles broaden the Paris Principles from covering primarily textual works to covering all types of materials, and in addition to choice and form of entry covered by the Paris Principles, these new principles cover all aspects of bibliographic and authority records used in library catalogs. It also provides "guiding rules that should be included in cataloguing codes internationally, as well as guidance on search and retrieval capabilities. [The] 2016 edition takes into consideration new categories of users, the open access environment, the interoperability and the accessibility of data, features of discovery tools and the significant change of user behaviour in general."[28] The principles are built on the conceptual models of FRBR and FRAD in addition to cataloging traditions from the past (e.g., Cutter, Ranganathan, Lubetzky).

The principles echo FRBR's user tasks in the "objectives and functions" of the catalog: to *find* resources in a collection, to *identify* a resource or agent, to *select* a resource that is appropriate, to *obtain* access to a resource, and to *navigate* and explore a catalog. The ICP states that the "highest principle for the construction of cataloguing codes should be the convenience of the users of the catalogue."[29] The general principles are enumerated in Textbox 7.1 on page 313.

The most extensive section of the ICP is called "Access Points." It notes that access points may be controlled or uncontrolled, the latter being such things as the *title proper* or *keywords* found anywhere in a bibliographic record. Names, titles, and other metadata needed for consistency in locating resources should be normalized, with variant forms used as references. The principles discuss choice of access points to include in a bibliographic description. They suggest that the titles of works and expressions should be controlled, as well as the names of the creators of works. The ICP also states that corporate bodies may be considered to be creators if works are expressions of the collective thought or activity of the body or if the wording of the title, in combination with the nature of the work, clearly indicates that the corporate body is responsible for the content. "Additional authorized access points for persons, families, corporate bodies, and subjects should be provided to bibliographic data, when deemed important for finding and identifying the bibliographic resource being described."[30]

Section 5.3 of the principles addresses authorized access points (AAPs). It begins with some general principles on the use of standards, on how to address different language and scripts, and on transliteration. Following those, there are principles that pertain to the choice and form of AAPs for various entities. These include the following:[31]

- Names (and the forms of those names) used in AAPs are to be represented consistently. The preferred names should be the ones found predominantly on manifestations or should be well-accepted conventional names.
- If more than one name is used by an entity, then one must be chosen as the basis of the AAP. If the variations are not representing different personas, then preference is given first to a commonly

known or conventional name, but finally to an official name if a commonly known or conventional name is not found.

- Successive different names for a corporate body that are more than minor variations are treated as names of new entities connected with references to earlier/later names.
- If needed, further identifying characteristics should be added to distinguish the entity from others with the same name.
- In all cases, variant forms not selected as the AAP should be included in the authority data as references or alternate display forms.
- Names of persons or families that consist of several words should be arranged in the order that follows the conventions of the country and language that is most associated with that person or family.
- Corporate names should be given in direct order unless they are part of a jurisdiction or the name implies subordination, in which case the authorized form should begin with the name of the jurisdiction or the superior body.
- A name for a work/expression may be a title that can stand alone, a name/title combination, or a title with qualifiers added. The preferred work/expression title should be the title found in the first manifestation of the work in the original language, or second, the title most commonly used.

When the creators of RDA began the process of developing a new content standard, they affirmed the role of the ICP as the basis for the cataloging principles used throughout RDA.[32] RDA has not only embraced the principles but also clearly follows the general guidelines for description, access points, and authority control. The following section addresses RDA's approaches to access and authority control; RDA's approach to bibliographic description is addressed in Chapter 7.

RDA: Resource Description & Access

The latest content standard intended for use in libraries, *RDA: Resource Description & Access*, not only provides guidelines for the description of resources (based on the WEMI entities) but also for the following activities:

- describing and establishing access points for agents (persons, families, and corporate bodies, including jurisdictions and conferences) that may act as creators, contributors, others associated with resources, and subjects of works;
- describing and establishing access points for works and expressions that may be contained in resources;
- describing relationships between a variety of entities; and
- assigning access points and relationship designators to bibliographic descriptions.

Each of these activities is addressed in the sections that follow.

Describing and Establishing Access Points for Agents

RDA, at the time of this writing (i.e., in its pre-LRM implementation), contains four chapters for describing the agents responsible for creating, contributing to, disseminating, or owning resources. In earlier content standards (e.g., AACR2), the focus was typically on establishing headings for persons, geographic jurisdictions, and corporate bodies, as well as uniform titles for works and expressions. In RDA the emphasis goes beyond establishing authorized access points; the focus is now on describing entities more robustly with a wide variety of attributes.

RDA provides general guidelines for describing agents. These guidelines address, among other topics, definitions, objectives and principles, languages and scripts, capitalization, accents and other diacritical marks, spacing of initials and acronyms, abbreviations, and the construction of authorized and variant access points. The general guidelines apply to each of the chapters that address different kinds of agents (i.e., persons, families, and corporate bodies).

The following section examines how to describe *persons* because it is a good illustration of the overall authority work process. Although not covered in this section, there are different and additional considerations, unique to corporate bodies, families, and places, that must be taken into account when describing those types of entities. For a complete overview of describing agents, please consult RDA directly.

Describing Persons: Names. In RDA's guidelines for describing persons, the initial focus is on establishing a *preferred name for the person*, which is the basis for an AAP and reflects the name most commonly associated with the person. The preferred name could be a person's real name, a pseudonym, a stage name, a title of nobility, a nickname, initials, or a word, phrase, or other appellation by which a person is known. Some persons use one name for their whole lives, but sometimes they may use different forms of that name. Thus, there are instructions on choosing the preferred name for a person when different forms of the same name have been used. For example, there may be variations in fullness, such as:

Michèle Valerie Cloonan versus **Michèle Cloonan**

The general rule for variations in fullness is to use the form most commonly found, and if that is not possible, to use the latest form of name, or if that is not clear, to use the fullest form of name. There may also be language variations:

Joan of Arc versus **Juana de Arco**

The general rule for forms in different languages is to use the form commonly found in most resources. If that is not clear, use the form of name most commonly found in reference sources from the person's home country. Additional rules are given for more complex scenarios. There also may be situations where the same name appears in different scripts:

Peter Ilich Tchaikovsky versus **Петр Ильич Чайковский**

In general, if the name appears in a script that is not used by the agency creating the metadata, then the name should be transliterated to the agency's preferred script, or, as an alternative, another form commonly found in reference sources may be used. Spelling variations may also be encountered:

Joanne A. Quitmeyer versus **Joann Quitmeyer**

If two or more different spellings of the same name are encountered, the general rule is to use the form of name that appears in the resource that was received first.

Some persons may use different names during their lifetimes. Sometimes, persons use real names in everyday life, but publish their works under different names; some use separate names for different bibliographic identities (e.g., different personas to write in different genres); and some change their names at various points in their lives. There are instructions for these situations as well. In general, choose the most well-known name for the person; if this is unclear, choose, in this order,

(1) the most frequently used name in resources,

(2) the name appearing most frequently in references sources, or

(3) the latest name used.

For names that have changed, use the latest name adopted by the person (e.g., use **Arlene G. Taylor** not **Arlene Taylor Dowell**). For those with more than one bibliographic identity, establish each identity separately using the name associated with it (e.g., use **Charles L. Dodgson** for his works on mathematics, but **Lewis Carroll** for his literary works).

Numerous instructions follow on how to record the name chosen as the preferred form. The guidelines are divided into instructions for surnames (including compound surnames, persons known by a surname only, hyphenated surnames, married persons known by a partner's name, and more), names containing titles of nobility, patronymics, given names only, names comprising words or phrases or initials, and others. In addition, there are instructions for recording variant names. The general rule for recording variant names is very simple. It states that the cataloger should record variant names when they differ from the preferred name. Specific instructions are then provided for real names, names in religion, secular names, earlier or later names of a person, alternative linguistic forms of names, and other alternative monikers.

Creating Access Points for Persons. After a preferred name is determined, an AAP can be established. An AAP is the standardized string of characters used to represent a person, place, corporate body, subject, or family in metadata. Although many believe the use of textual character strings is an antiquated approach to identity management (and it is), it is still the most human-friendly approach available at this time. In the future, more machine-processable approaches may be employed (e.g., the use of URIs), but for now catalogers establish a unique AAP for the entity and use it consistently in all metadata associated with that entity in order to allow for collocation in retrieval.

When constructing an access point, catalogers begin with the preferred name chosen for the person. Most names contain multiple parts, so the cataloger must determine the entry element. The entry element, the part of the name entered first, should be recorded according to the conventions used with the person's native language or in the person's country of origin or residence. In many parts of the world that means the first element in the AAP should be the family name (also referred to as *surname* or *last name*), followed by a comma and the rest of the name. For example, in the United States, the name *Rebecca Green* would be entered as **Green, Rebecca**. In Spain, where persons have compound surnames (i.e., two family names), a person is entered under the first of two family names. For example, librarian Jesús Alonso Regalado's preferred name would be established as **Alonso Regalado, Jesús**. In other parts of the world, surnames are less common or less impor-tant and persons may be entered under their given names instead (e.g., **Björk Ingimundardóttir**). If the preference of the person is known or discernible, that preference should take precedence over standard practices.

When the preferred name is unique, it can stand alone as an AAP, but many names are rather common. In these cases, additional data elements are required to identify the entity precisely. The following are the most typical additions to a name:

- middle initials
- fuller forms of name
- birth and death dates
- periods of activity
- titles associated with the person (e.g., royal titles, titles of nobility, religious titles, *Saint*, *Spirit*)
- professions or occupations (e.g., Notary, Composer, Veterinarian)
- terms of rank, terms of honor, terms of office (e.g., Captain, Ph.D.)
- other designations associated with the person

For example, there is more than one Rebecca Green in the world. The name is not unique, and would, therefore, need additions in order to formulate unique AAPs that differentiate the different Rebecca Greens. The following are the Rebecca Greens currently in the LCNAF:

- **Green, Rebecca, 1743–1806**
- **Green, Rebecca, 1952–**
- **Green, Rebecca, 1975–**
- **Green, Rebecca, 1978–**
- **Green, Rebecca, 1979–**
- **Green, Rebecca, 1980–**
- **Green, Rebecca E. (Rebecca Erin), 1972–**
- **Green, Rebecca J. (Rebecca Jane), 1949–**
- **Green, Rebecca L.**
- **Green, Rebecca L. (Rebecca Lynn)**
- **Green, Rebecca Louise, 1978–**
- **Green, Rebecca Meade, 1795–1867**
- **Green, Rebecca, RN**

Each Rebecca Green has an authority record. Each authority record contains the established AAP, any variant forms of name used by that Rebecca Green, and possibly some additional identifying information.

Other Attributes for Describing Persons. The RDA metadata elements for describing persons are listed in the bullet points below.

- **Name of the Person**
- Date Associated with the Person, including **Date of Birth** and/or **Date of Death**
- **Title of the Person**, if applicable
- Fuller Form of Name
- **Other Designation Associated with the Person**, required for specific categories of persons
- Gender
- Place of Birth
- Place of Death
- Country Associated with the Person
- Place of Residence, etc.
- Address of the Person
- Affiliation
- Language of the Person
- Field of Activity of the Person
- **Profession or Occupation**, required for some unusual names
- Biographical Information
- **Identifier for the Person**

This list shows the most obvious contribution of the FRAD conceptual model to contemporary cataloging practices (nearly all the elements were originally enumerated in FRAD rather than in FRBR). Most of the elements are fairly self-explanatory. Some are identified in RDA as *core elements*; they are presented here in boldface type. Core elements are always recorded, if they are applicable to the entity being described and the information is found in sources. The rest of the elements are recorded in the authority data based on cataloger's judgment. At this time, most authority records in the LCNAF contain few of these elements for a variety of reasons. It may be because some records were created under earlier sets of cataloging guidelines; some records are less than complete because additional metadata was unavailable or not applicable; or, it could be that, in some cases, the cataloger simply chose not to do thorough authority work. An example of a brief personal name authority record (with labels added) is found in Figure 8.4.

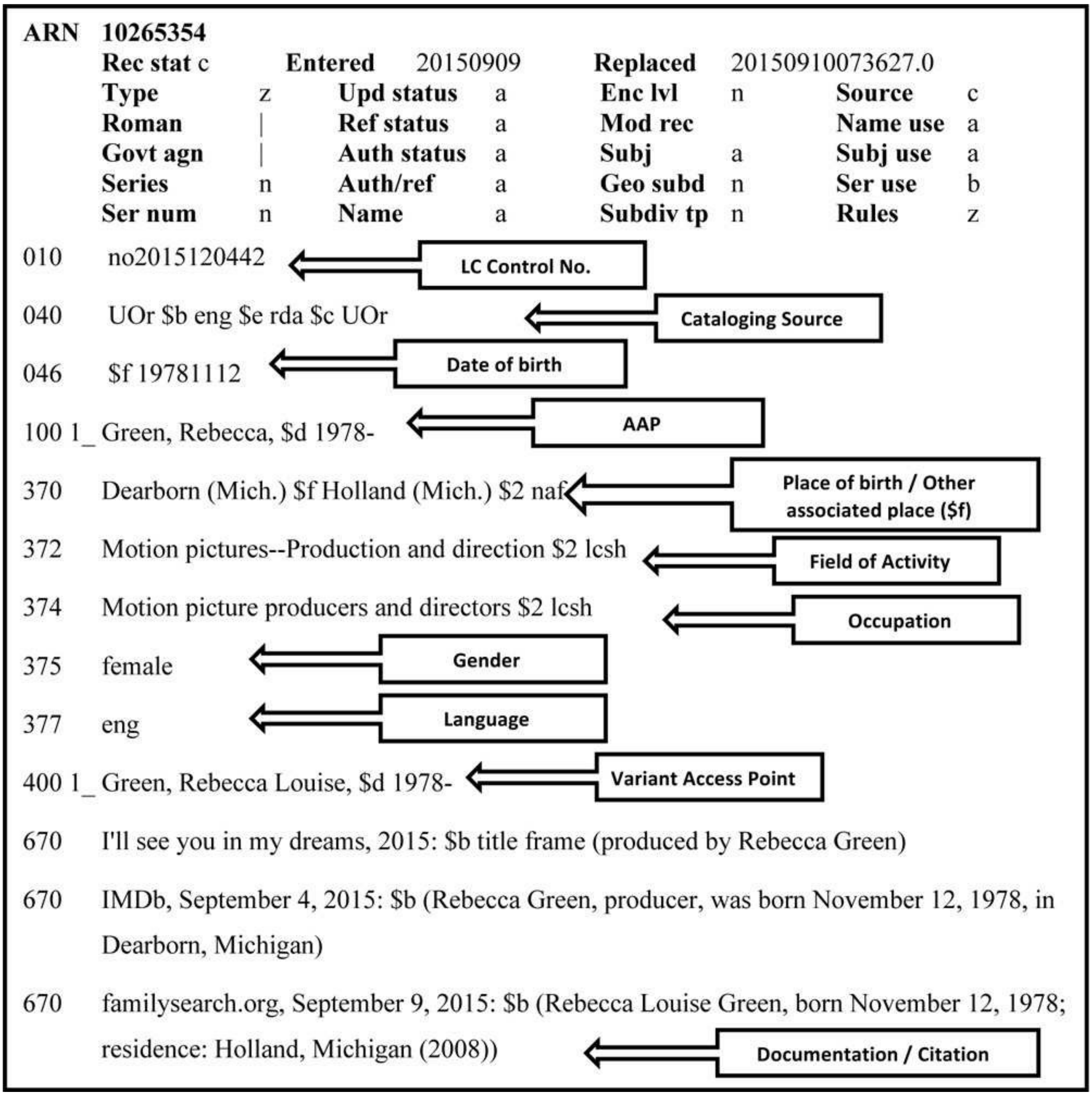


Figure 8.4. A Labeled Authority Record Example. (Source: OCLC Connexion, Authorities—Record Number 10265354)

Describing and Creating Access Points for Works and Expressions

Describing Works and Expressions. RDA not only offers instructions on how to describe agents but it also provides guidelines on how to record authority data for works and expressions (the abstract, intangible content-related entities described in FRBR and LRM). Why do catalogers create metadata about works and expressions? Generally, there are two reasons:

- Different expressions and different manifestations of a work are not always known by the same title, and
- Different works may be known by the same title.

The AAP is the primary mechanism by which catalogers unambiguously (1) represent a work or expression embodied in a resource, or (2) represent a relationship between one work or expression and another (i.e., a related work). By creating a unique AAP, various resources containing the same intellectual or artistic content can be collocated in the catalog, no matter their individual titles proper. And resources with identical titles but with different content can be differentiated through the use of an AAP. AAPs not only bring together the like, they separate the unlike, or, more accurately, AAPs collocate entities that are the same and separate entities that are different (also referred to as *disambiguation*).

A favorite example to illustrate the necessity for authority data for works and expressions is William Shakespeare's *Hamlet*. This work has been known by different titles over the years—across various languages, scripts, and transliterations. The following are just a few of the variations that can be found:

- *Hamlet*,
- *Hamlet, Prinz von Dännemark*,
- *Gamlet*,
- *Khamlėt*,
- *Shakespeare's Hamlet*,
- *Tragedien om Hamlet, Prins av Danmark*,
- *The Tragedy of Hamlet*, and
- *The Tragicall Historie of Hamlet, Prince of Denmarke*.

All these resources are brought together in the catalog under a single preferred title, *Hamlet*, which when combined with the authorized access point for Shakespeare provides a standard citation for the work: **Shakespeare, William, 1564–1616. Hamlet**.

Establishing an AAP is usually the most pressing reason for creating authority data for a work or an expression, but metadata other than the name/title AAP is also recorded. There are a dozen work-related and expression-related elements that may be recorded; two of which are considered core in RDA. Four additional elements are considered core *if* a preferred title is not unique and needs disambiguation. In addition, there are specific instructions and elements aimed at establishing authority data for musical works, legal works, treaties, religious works, and official communications; each may require one or more core elements to be recorded.

As is the case with establishing names for agents, establishing titles for works and expressions entails choosing one title to be the preferred one, while the others are recorded as variant titles. The primary consideration in choosing the preferred title is the creation date. For works created *after 1500*, catalogers are to choose the best-known title in the original language as the preferred title. For works created *before 1501*, catalogers are to choose the best-known title in the original language that is found in reference sources. If those instructions cannot be easily followed, alternative approaches are provided. There are specific instructions for special cases (e.g., how to record a preferred title for parts of a work) and guidelines for recording variant titles as well.

Establishing Access Points for Works and Expressions. With all the metadata about a work or expression collected, the AAP can be established. There are sets of guidelines for establishing AAPs for five different types of resources: (1) general works and expressions, (2) musical works, (3) legal works, (4) religious works, and (5) official communications. The general instructions can also be used to create authority data for series (a series can be considered a work in its own right). In this section, we examine some aspects related to general works and expressions, but we do not explore the four more specific types of works. As is the case with all aspects of cataloging, please see other resources, such as RDA itself and Daniel N. Joudrey, Arlene G. Taylor, and David P. Miller's *Introduction to Cataloging and Classification* for more detailed information.

Authorized access points for works come in two basic forms:

(1) *name/title* access points

(2) *title-only* access points

The first is used when there is an identifiable creator(s) for the work; the second is used for anonymously created and a few other types of works. Determining the agent most responsible for creating a work is a longstanding practice in cataloging. Until the advent of RDA, this was known as choosing the *main entry*. This terminology has been omitted from RDA, but the basic concept has lived on as the first step in establishing a name/title AAP to identify a work.

The basic RDA guidelines for establishing an AAP are organized around several conditions of creation:

- **Works created by one agent**: The AAP for the work comprises the AAP for the agent and the preferred title for the work (e.g., **Cassatt, Mary, 1844–1926. Children playing with a dog**).

- **Collaborative works**: The AAP for the work comprises the AAP for the agent with principal responsibility for the work and the preferred title for the work (e.g., **Bernstein, Leonard, 1918–1990. Candide**). If the agent with principal responsibility is not clear, the first-named agent is used.
- **Compilations of works by different agents**: When compilations are composed of contributions by different creators, no single agent is chosen as the primary creator; instead, the AAP for the work is title only (e.g., **Norton anthology of world masterpieces**).
- **Adaptations and revisions**: If the work is presented as a new version (i.e., expression) of a work, the AAP is identical to that of the original expression of the work. If the resource is a significant transformation of the original content, the agent responsible for the new content is the name featured in the AAP (e.g., due to large-scale revisions, the fourth edition of a particular work is known as **Chan, Lois Mai. Immroth's Guide to the Library of Congress classification**, whereas the first through third editions are referred to as **Immroth, John Phillip. A guide to Library of Congress classification**).
- **Commentary, Annotations, Illustrative content, etc., added to a previously existing work**: If the commentary, annotations, or illustrations are the focus of the resource, then an AAP reflecting these features is created. For example, if the focus of a resource were on the commentary rather than on the original text of Shakespeare's *Hamlet*, then the AAP would be created for the commentary (e.g., **Mategrano, Terri. Shakespeare's Hamlet**). If the resource is presented as little more than a new edition of a previously existing work with some extra content, then the AAP is identical to the original work (e.g., the established AAP—**Carroll, Lewis, 1832–1898. Alice's adventures in Wonderland**—is used for the original edition as well as for the 150th anniversary edition featuring illustrations by Salvador Dalí).
- **Works of uncertain or unknown origin**: If there is no agreement among scholars as to the agent responsible for creating a work or the creator is truly unknown, then only the title is used as the AAP (e.g., **Beowulf**). If reference sources generally concur on a probable creator, that creator's name is used as the beginning part of the AAP for the work (e.g., **Aneirin. Gododdin**).

There are, of course, some exceptions, options, and alternatives for some specific resource types and conditions.

Although in many cases there is very little difficulty in determining the agent primarily responsible for the creation of the work, for some resources it can be challenging. Some resource types, such as sacred scripture, compilations, motion pictures, and other aggregate works, typically do not have a single identifiable creator. These types of work are usually known by title only.

There are special instructions for certain types of material (musical works, legal works, etc.) to address some difficulties in determining the creator. For example, works created by corporate bodies may seem challenging at first. RDA, however, gives guidance for this with its listing of eight categories into which a resource must fall in order for a corporate body to be considered its creator. The list includes such groupings as works about the administration of the body itself, works that reflect the collective thought of the body, hearings conducted by the body, reports of a conference or other such event, and works of art or musical performances created by two or more artists/musicians acting as a group. If a work is created by a corporate body, but does not fall into one of the categories, the AAP is title only, *if* there is no person or family also involved in its creation.

In some cases, no matter the type of creator, additional data elements may be added to make the AAP unique. These include the elements listed below:

Element	Example
Form of work	**Gale, Zona, 1874–1938. Miss Lulu Bett (Novel)**
Date of work	**Wilson, Lanford, 1937–2011. Works. 1996**[33]
Place of origin	**Maryland gazette (Annapolis, Md. : 1727)**
Other distinguishing characteristics	**Bulletin (Alaska State Council on the Arts)**

Particular expressions of a work also can have AAPs established for them. For each expression, the AAP for the work is duplicated and additional data elements are added to the AAP to distinguish that particular expression—perhaps using language, content type, date, another distinguishing characteristic, or a combination of these, if needed.

AAP for work:	**Shakespeare, William, 1564–1616. Hamlet.**
AAP for expression:	**Shakespeare, William, 1564–1616. Hamlet. Romanian.**

As with other types of entities, variant access points may be recorded in the authority data. For example, in the authority record for **Shakespeare, William, 1564–1616. Hamlet** in the LCNAF, there are variant access points for the variant titles *The Tragedy of Hamlet* and *Shakespeare's Hamlet*.

Describing Relationships

In addition to describing entities, RDA provides instructions for connecting entities through relationships. There are guidelines for a variety of relationship types, including those describing

- relationships between agents and resources,
- relationships among works, expressions, manifestations, and items (e.g., related works, related manifestations), and
- relationships among agents (e.g., person-to-family relationships, person-to-person relationships).

RDA uses four different approaches to recording data; this is referred to as the *four-fold path*, which is described in Chapter 7. Any of the methods in the four-fold path can be used to describe relationships between entities, if deemed appropriate. The four methods include: unstructured descriptions, structured descriptions, identifiers, and IRIs. In authority data, relationship designators may be used with any of the four paths. A *relationship designator* is a device (i.e., a label, phrase, or term) that clearly identifies the specific nature of the relationship that exists between entities; it provides additional context. RDA offers, in its appendices, several long lists of relationship designators based on various types of entities. The lists contain two reciprocal forms of each relationship type so that the relationships can be seen from either direction (e.g., contained in / container of). A small sample of designators is provided in the following table:

Type of Relationship	Examples
Relationships between WEMI and agents	**author; book artist; calligrapher; cartographer; composer; designer; editor; enacting jurisdiction; interviewer; plaintiff; producer; researcher; sponsoring body; translator; visual effects provider; writer of foreword; and so on**
Related works, expressions, manifestations, and items	**based on (work); contained in (work); container of (work); digest of (work); electronic reproduction (manifestation); facsimile of (item); guide to (expression); illustrations for (work); mirror site; prequel to; sequel; revised as; and so on.**
Related persons, families, and corporate bodies	**absorbed corporate body; alternate identity; colleague; family member; founder; graduate; hierarchical subordinate; real identity; sponsoring family; teacher; and so on.**

In the authority data established for *The Oresteia*, a trilogy of plays by Aeschylus, one can see relationship designators in use:

AAP:	**Aeschylus. Oresteia**
See also:	Container of (work): **Aeschylus. Agamemnon**
See also:	Container of (work): **Aeschylus. Choephori**
See also:	Container of (work): **Aeschylus. Eumenides**

The relationship designator *container of (work)* indicates that *The Oresteia* contains the three separate works listed in the data. The reciprocal relationship is found in the authority data created for each of the three individual plays:

AAP:	**Aeschylus. Agamemnon**
See also:	Contained in (work): **Aeschylus. Oresteia**

Another example of the types of relationship recorded in authority data is associated with agents. A person might use separate identities when creating different sorts of works. Earlier, we mentioned **Charles L. Dodgson**, who wrote about mathematics, and his alternate identity, **Lewis Carroll**, who wrote literary works. Sometimes the situation is even simpler: some authors write under more than one name for whatever reasons they choose, with no concerns about establishing different bibliographic identities. No matter how this comes about, the creators of authority data try to connect different names and identities/personas used by the same actual person in their authority data. For example, one finds this information in Isaac Asimov's authority record:

AAP:	**Asimov, Isaac, 1920–1992**
See also:	**French, Paul, 1920–1992**
See also:	**Dr. "A", 1920–1992**

Authority data is the result of catalogers gathering together information about various entities that are part of the bibliographic universe. It helps catalogers better understand the agents that are responsible for the creation and enhancement of works and expressions. It also helps catalogers better understand the content and the origins of works (and their varying expressions) that appear in resources. In addition, authority data provides catalogers with unique identifiers and AAPs for use in bibliographic records. AAPs in bibliographic records allow for collocation and separation. A standardized name, title, or name/title brings together all the resources

- by a particular creator,
- containing a particular work or expression,
- on a particular subject, or
- in the same genre/form.

In other words, AAPs assist in meeting Cutter's objects of a catalog (described in Chapter 3) and the IFLA user tasks (described in Chapter 5). AAPs, with relationship designators attached, provide users with a greater understanding of the relationships among entities and the roles that various agents play in the development of resources.

The Use of Access Points in Bibliographic Records

Access points represent relationships in the bibliographic universe. They are assigned to bring out names, titles, and subjects associated with resources. They are often found on the preferred sources of information for the resource (e.g., a title page for a book). If the preferred sources of information prove to be inadequate, other parts of the resource may be examined. Outside sources (e.g., reference works) may also be consulted as needed.

Typically, *name* access points are AAPs of the agents associated with a resource. Each agent has some type of relationship to the resource, which may be identified through a relationship designator (e.g., author, artist, choreographer) and, when MARC is being used, the tag for the field (e.g., 1XX for a primary creator; 7XX for subsequent creators and contributors). *Title* access points include both controlled and uncontrolled titles. A *controlled title* is one that has been established as part of authority work; it is used for work titles (including series) and expression titles. An *uncontrolled title* is the title that appears on a manifestation; it is a transcribed title proper. It can also include variant forms of a title proper (e.g., spine title, cover title, corrected title proper). Although *title proper* is an attribute for a resource, it can also be viewed as a relationship between the resource and the name(s) that the resource has been given and the work(s) and expression(s) that the resource contains. Controlled *subject* access points also represent relationships; for example, a subject heading reflects the relationship between a work and a res, the thing(s) the work is about. These relationships—these access points—make resources findable in the catalog in a consistent manner and provide collocation of related materials. In the future, it is likely that each AAP will be replaced by or at least supplemented with a standard identifier or an IRI rather than relying on the structured text string alone.

At the time of this writing, RDA identifies three core types of agent relationships that may be represented by access points:

- Creator
- Contributor
- Other Agent Associated with Work

RDA does list other possible relationships between agents and the WEMI entities, but these relationships are usually not, at this time, identified through controlled access points. These roles include publishers, producers, manufacturers, distributors, owners, custodians, and others that may be related to manifestations and items in some other fashion. In a typical bibliographic description today, this type of data is transcribed as an attribute (e.g., publisher's name) or is recorded as a note rather than being treated as an access point. Access points tend to focus on the entities that are considered *bibliographically significant*—those entities that made important contributions to the intellectual and artistic content of a resource, as opposed to the physical production or dissemination of the resource. This, however, may change in the future.

Unlike previous sets of cataloging rules, RDA places no real restrictions on the number or kinds of access points a cataloger may add to the bibliographic description. However, only a primary or first-named creator, a title proper, a work title (if applicable), and one subject are considered *core*; the inclusion of all other access points is based on cataloger's judgment. This means that different catalogers could end up with very different looking metadata when describing the same resource. On the other hand, a conscientious cataloger is expected to include as much metadata as needed to ensure that the resource is findable, identifiable, understandable, and so on. In fact, one of the things that catalogers have liked about RDA is that it allows for more than three access points for agents of the same type to be included in the metadata (AACR2 restricted the number of access points added). That does not mean, however, that all catalogers will record all possible access points; some catalogers may be more selective in their access points. Nor does it mean that catalogers will stick with just the core requirements (remember, *core* refers to the bare minimum). Cataloger's judgment and/or local institutional policies can play a role in determining how many access points are included.

Creators can be persons, families, or corporate bodies. As discussed above, in the section on establishing access points for works and expressions, one must think about the conditions of creation. If, for example, only one agent created the work, the situation is fairly easy. Only one access point for a creator is needed. When more than one agent has played a role in the creation (or subsequent revisions) of a work, then their respective activities must be considered. When two or more agents have played the same role in the creation of the work, all may be recorded as access points, but one will be identified for use in the AAP established for the work. For the resource, *Introduction to Cataloging and Classification*, 11th ed., by Joudrey, Taylor, and Miller, you would include the following access points:

Creator:	**Joudrey, Daniel N.**, author.
Creator:	**Taylor, Arlene G., 1941–** author.
Creator:	**Miller, David P. (David Peter), 1955–** author.

All three creators are included in the metadata (with relationship designators added at the end of the AAP). When it comes to establishing the AAP for the work, however, only one would be selected:

AAP for the work:	**Joudrey, Daniel N. Introduction to Cataloging and Classification.**

After all creators (or selected creators) have been added to the description, catalogers must consider if there are contributors that have played significant roles in developing the resource. For example, editors, illustrators, translators, performers, and so on, may be included as contributors. *Contributor* is not considered a core element in RDA, but a conscientious cataloger is not going to omit key contributors just because they are not required.

Once the agents responsible for the resource have been identified and included as access points, the cataloger considers additional access points for the description. These include titles proper, work titles, series titles, related works or expressions (name/title access points), and variant titles (including cover titles, caption titles, running titles, titles on containers, title-bar titles, corrected titles, spine titles, and parallel titles). Subjects, as stated earlier, are also access points and are discussed in later chapters. The inclusion of access points makes resources findable in the catalog in a consistent and predictable manner.

For an example of access points included in a MARC bibliographic record, please see Figure 6.3 on page 258. Personal name access points are found in the 100 and the first two of the 700 fields. The title proper access point appears in subfield *a* ($a) of the 245 field. Subject access points are in the four 6XX fields. The series access point is in the 830 field, and a name/title access point for a related work appears in the third 700 field.

Anglo-American Cataloguing Rules, Second Edition (AACR2)

One might ask, "If RDA has replaced AACR2, why is the earlier content standard still being addressed in this book?" As we stated in Chapter 7, at this time, not all North American libraries have switched to RDA.

For a variety of reasons, such as financial concerns, philosophical differences, and/or simple intransigence, some libraries have chosen to continue using AACR2 rather than make the change. In addition, a brief introduction to AACR2 may help new LIS professionals understand legacy data they may encounter.

Prior to the release of RDA, the second edition of the *Anglo-American Cataloguing Rules* was the content standard in use in North America (and beyond) for over three decades. The principles that underlie access points in AACR2 are based upon the Paris Principles, which were accepted internationally at the International Conference on Cataloguing Principles, held in Paris in 1961.[34] Names, titles, and name/title combinations are the candidates for access points. AACR2 calls for access points for persons, corporate bodies, geographic names, and titles, as well as combinations of names and titles. Some specific access points called for by AACR2 include collaborators, editors and compilers, corporate bodies, titles (including titles of series), related works, and sometimes translators and illustrators. These are fairly book-oriented choices, reflecting the origins of AACR2, whereas in fact, many access points are for performers, choreographers, programmers, and cartographers, among others.

Main Entry, or, Primary Access Point

According to the Paris Principles, one of the access points is selected to be the primary access point. The Paris Principles called this the *main entry* and said that it must be a *full entry*. The concept of a full entry made sense in the days of print catalogs (i.e., card, book, and microform catalogs) when it was necessary to place as many copies of the same surrogate record into the catalog as there were access points. (See Figure 8.1 for an example of a title card.) In the days of handwritten and typed catalog entries, and later in book catalogs especially, it was often necessary to save time and space by having only one copy of the surrogate record created as a full entry (i.e., containing a complete set of all elements of the record as provided by the cataloger). The rest of the entries were abbreviated. The Paris Principles stated that the place where the full entry should be filed is the alphabetical location of the primary access point or main entry. Soon it became customary to refer to the full entry as the *main entry*.

Use of *main entry* to mean both *full entry* and *primary access point* caused confusion throughout the last two decades of the twentieth century. Catalogs had long since given up the practice of having one full entry and several abbreviated copies of each surrogate record. Instead, as technology advanced, every entry became a duplicate of the full entry. The difference among the copies was that each copy had as its top line one of the chosen access points. The concept of full entry is not applicable in online catalogs.

Main Entry Controversy. Controversy over main entry (in the sense of primary access point) is four decades old. In 1978 Seymour Lubetzky and Michael Gorman disagreed quite publicly at a conference on the emergence of AACR2.[35] AACR2's introduction included the observation that the concept of main entry might have outlived its time. However, there had not been enough time to research the idea, and so the concept was included in AACR2. It stated that a library that wished to could choose access points without designating one as the main one. Lubetzky, whose work underlies the Paris Principles, argued that this was a step backward instead of forward. Gorman believed that the time had come to make the distinction only when necessary to cite one work on the record for another work, a situation that he guessed would happen "less than 1 percent" of the time.[36] Research, however, has shown this estimate to be quite low. Richard Smiraglia found that approximately 50 percent of a sample of works had derivative relationships.[37] As explained earlier, derivative relationships exist when one work is derived from another (e.g., editions, translations, adaptations, performances). A bibliographic description for a resource containing a derivative work should cite or refer to the work from which it has been derived.

Justification for Choice of Main Entry. In Chapter 3, Cutter's objects of a catalog were quoted. Paraphrased, they are

- to enable a person to find a book when author, title, or subject is known;
- to show what the library has by a given author, on a given subject, or in a given kind of literature;
- to assist in the choice of a book as to its edition or as to its character.

In many ways, choosing a main entry was seen as a method to carry out Cutter's objectives. Reasons given for a need for a main entry include the following:

- to provide a standard citation form in order to show relationships among resources, identify works about other works, and identify works contained in larger works

- to provide sub-arrangement under subjects (and as a corollary, to provide a way to sub-arrange items under classification numbers) in order to show prolific creators in a field and bring together manifestations of a work
- to provide for collocation in the catalog of all manifestations containing the same work, even though the resources may be published with different titles (e.g., translations), or editions of a work may be written by different authors
- to assure a judgment of the most important and most predictable access point in situations where reduction of cataloging time has been mandated.

In a catalog record, a main entry is not enough to ensure that users will be able to retrieve the information they need. Additional access points, called *added entries* in AACR2, supplement the primary access point.

Establishing Headings

In addition to the rules for choosing access points (i.e., the main and added entries), AACR2 also provides instructions for creating authority-controlled headings for persons, corporate bodies, and other entities. AACR2 follows certain principles in its rules for making these decisions. These principles are presented in Chapters 22–25 of AACR2. Although the terminology is different, the processes for determining headings are quite similar to those for creating access points for agents and titles, explained for RDA above.

For a comparison of some major differences between AACR2 and RDA, see Table 8.1. For more detail about how to choose access points and establish headings and uniform titles, please see AACR2, which is available in the RDA Toolkit. Or, for a guide to AACR2 instructions, please consult Arlene G. Taylor's tenth edition of *Introduction to Cataloging and Classification*, published in 2006.[38]

Table 8.1. Key Differences between RDA and AACR2 with Respect to Access Points.

AACR2	RDA
The term *author* is often used for those responsible for the creation of works.	When identifying those responsible for the creation of works, the more neutral term *creator* is used instead.
The term *heading* is used to refer to the standardized name established for an entity.	An *authorized access point* (AAP) is established for entities under the instructions in RDA.
Under the rules in AACR2, catalogers choose a *main entry* and a limited number of *added entries* as access points for the description.	In RDA, catalogers choose the *agents* responsible for a resource (e.g., creators and contributors). One agent is identified as the *primary* creator of the work (or the first-named, if primary responsibility is unclear or shared), when creating a name/title access point to cite a work.
The *rule of three* limits the number of access points that may be added to a description.	As many access points as deemed necessary may be added to the bibliographic description.
There are no rules for families as creators.	Families can be considered creators of resources.
The relationship between an access point and the resource may be unclear.	Relationship designators (e.g., narrator, illustrator) may be applied to the access point to clarify the role played by the agent.
When a compilation of individual works (e.g., short stories) lacks a collective title, the resource is entered under the author of the first work.	A compilation is usually entered in the catalog under its title, but if there is no collective title the cataloger constructs separate access points for each of the works in the compilation.
Authority work primarily consists of establishing an authorized name for an entity and its cross-references (with some additional data that may be needed to resolve conflicts).	Under RDA, besides establishing a name and cross-references, there are many additional elements about an entity that may be recorded (e.g., occupation, birth place).

Metadata Authority Description Schema (MADS)

MADS is an XML schema for an authority element set that may be used to provide metadata about agents (people, organizations), events, and terms (topics, geographic names, genres, etc.).[39] MADS was created to serve as a companion to the Metadata Object Description Schema (MODS) described in Chapter 7. It is intended to be a simpler way to encode authority data than the MARC 21 Authority Format, while still remaining compatible with MARC-based authority data.

Each MADS record must contain the schema's main element, called **<authority>**, that contains the authoritative form of the name or term covered by the record. The record may have any additional number of elements labeled **<related>** and/or **<variant>**. Elements marked **<variant>** are references from an alternative name or term to the authorized access point (e.g., from Knowles, Beyoncé to **Beyoncé, 1981–**). Elements marked **<related>** are references from other authorized forms of name or term to the authorized access point (e.g., from **Dodgson, Charles Lutwidge, 1832–1898** to **Carroll, Lewis, 1832–1898**).

Each of these top-level elements (**<authority>**,**<related>**, and **<variant>**) can have one or more of the following descriptor elements:

- name
- titleInfo
- topic
- temporal
- geographic
- hierarchicalGeographic
- genre
- occupation

The following are additional miscellaneous elements that might be added to a MADS record:

- affiliation
- classification
- extension
- fieldOfActivity
- identifier
- language
- note
- recordInfo (i.e., the metadata for the record itself)
- URL

A MADS record can be as basic or extensive as required and can be customized according to the level of description desired for the type of system in which it is intended to be used. An example of an excerpt from a simple MADS record can be viewed in Figure 8.5.

STANDARDS FOR ARCHIVES

Although the archival community has long recognized the need for controlled access points, it has been relatively recently that they have created their own standards. The standards addressed in this section are

```xml
<mads:mads xmlns:mads="http://www.loc.gov/mads/v2"
xmlns:xlink="http://www.w3.org/ 1999/xlink"
xmlns:mods="http://www.loc.gov/mods/v3"  xmlns:xsi="http://www.w3.org/
2001/XMLSchema-instance"  xsi:schemaLocation="http://www.loc.gov/mads/
http://www.loc.gov/standards/mads/mads.xsd http://www.loc.gov/mods/v3
http://www.loc.gov/standards/mods/v3/mods-3-2.xsd">
    <mads:authority lang="eng">
        <mads:name type="personal" authority="naf" >
            <mads:namePart>Abbott, Elizabeth</mads:namePart>
            <mads:namePart type="date">1942- </mads:namePart>
        </mads:name>
    </mads:authority>
    <mads:variant type="other" lang="eng">
        <mads:name type="personal">
            <mads:namePart>Namphy, Elizabeth Abbott</mads:namePart>
            <mads:namePart type="date">1942- </mads:namePart>
        </mads:name>
    </mads:variant>
    <mads:note type="source">Tropical obsession, 1986: title page (Elizabeth Abbott
        Namphy)</mads:note>
     <mads:note type="source">Penguin.ca website, viewed 25 June 2014: author page
        (Elizabeth Abbott is the former Dean of Women at the University of Toronto and
        the bestselling author of A History of Celibacy and A History of
        Mistresses)</mads:note>
    <mads:identifier type="lccn">nb2014014243</mads:identifier>
    <mads:fieldOfActivity>Authorship</mads:fieldOfActivity>
    <mads:fieldOfActivity>History</mads:fieldOfActivity>
    <mads:extension>
        <mads:gender>
            <mads:genderTerm>female</mads:genderTerm>
        </mads:gender>
    </mads:extension>
    <mads:extension>
        <mads:profession>
            <mads:professionTerm>Authors</mads:professionTerm>
            <mads:professionTerm>Historians</mads:professionTerm>
        </mads:profession>
    </mads:extension>
    <mads:recordInfo>
        ....
    </mads:recordInfo>
</mads:mads>
```

Figure 8.5. An Excerpt from a Simple MADS Record, Based on LC MARC Authority Record nb2014014243. (MADS coding shown in bold)

- *International Standard Archival Authority Record for Corporate Bodies, Persons, and Families* (ISAAR(CPF))
- *Describing Archives: A Content Standard*(DACS)

- Encoded Archival Context—Corporate Bodies, Persons, and Families (EAC–CPF)

International Standard Archival Authority Record for Corporate Bodies, Persons, and Families (ISAAR(CPF))

In 1996 the International Council on Archives (ICA) published the *International Standard Archival Authority Record for Corporate Bodies, Persons, and Families* (ISAAR(CPF)), with a substantially revised second edition following in 2004.[40] ISAAR(CPF) provides guidance on creating archival authority records for corporate bodies, persons, and families who have created and maintained the records, documents, and other resources that comprise archives. The precept behind this work is that the creation of separate descriptions of creators of archival collections, which are then linked to descriptions of those collections (e.g., finding aids, web pages), provides an efficient means of capturing and managing the contextual information that is vital to the discovery, use, and understanding of archival collections. The second edition has increased consistency with related international standards such as the ICA standard for archival description, *ISAD(G) General International Standard Archival Description* (discussed in Chapter 7). The Encoded Archival Context–Corporate Bodies, Persons, and Families (EAC–CPF) standard for machine readable interchange of ISAAR(CPF)-compliant data presents a companion data structure standard. Part II of the 2nd edition of DACS is derived from the ISAAR(CPF) standard.

The bulk of ISAAR(CPF) consists of a listing of elements needed for an authority record. A purpose is given for each element, followed by the rule for transcribing it and examples of its application. Elements are divided into four areas:

- **Identity area**: includes type of entity; authorized form(s) of name; parallel forms of name; standardized forms of name according to other rules; other forms of name; and identifiers for corporate bodies
- **Description area**: includes dates of existence; history; places; legal status; functions, occupations, and activities; mandates/sources of authority; internal structures/genealogy; and general context
- **Relationships area**: includes names/identifiers of related corporate bodies, persons, or families; category of relationship; description of relationship; and dates of the relationship
- **Control area**: includes authority record identifier; institution identifiers; rules and/or conventions; status; level of detail; dates of creation, revision, or deletion; languages and scripts; sources; and maintenance notes

Relationships to resources are also accommodated in ISAAR(CPF). Chapter 6 of the standard outlines the type of information that should be included to describe adequately the relationships between agents and archival materials. The description of a relationship may include identifiers, titles, resource types, the nature of the relationship, and dates associated with those resources and/or relationships.

Describing Archives: A Content Standard (DACS)

DACS is a standard that has been accepted and used by the U.S. archival community in the way that AACR2 and RDA are used by the library community. DACS has two parts—one for describing archival materials and one for creating archival authority records. The first part is discussed in Chapter 7 of this text. The introduction to the second part emphasizes that in order to understand archival materials completely, one must have some knowledge of the context in which they were created. The three steps required for establishing such a context are[41]

- identifying the persons, families, and corporate bodies that played important roles in the creation of the materials;
- assembling biographical information about the individuals and families and collecting historical data about the corporate bodies, including their structure, functions, and relationships; and
- providing standardized forms of the names of the persons, families, and corporate bodies in order to facilitate retrieval.

Chapter 9 in DACS, "Archival Authority Records," covers the first of the steps just listed, and Chapter 10, "Form of Name," provides guidance on the use of companion standards for the third step. Chapters 11–14 describe the elements of archival authority records, the management of those records, and related archival materials and other resources.

Chapter 9 identifies the necessity to establish archival authority records, levels of description, and the minimum elements that need to be included in an authority record: the authorized form of name, the type of entity, dates of existence, and an authority record identifier.[42] In addition to guidelines for the authorized form of name, Chapter 10 provides guidelines for recording the type of entity, variant forms of names, and identifiers for corporate bodies (including the jurisdiction that assigned the identifier). Chapter 11 provides the descriptive elements for corporate bodies, persons, and families; Chapter 12 attends to relationships between the entity being described in the authority record and other corporate bodies, persons, and families; Chapter 13 lays out the information necessary for authority record management; and Chapter 14 deals with relationships between the entity being described in the authority record and archival materials and other resources. There is a clear map to the elements in the international standard, ISAAR(CPF).

Encoded Archival Context–Corporate Bodies, Persons, and Families (EAC–CPF)

Encoded Archival Context–Corporate Bodies, Persons, and Families (EAC–CPF),[43] a standard for encoding data in archival authority records, was released in 2010 and formally adopted by the Society of American Archivists in 2011. EAC–CPF, like EAD (described in Chapter 7), uses XML for encoding, but while EAD describes archival collections, EAC–CPF describes agents that may act as the creators of archival materials or that may be the subjects of archival collections. EAC–CPF records do more than just control all the different forms of name used by a person, corporate body, or family. According to Daniel Pitti, they also describe "their essential functions, activities, and characteristics, and the dates and places they were active."[44] Pitti goes on to explain that creator descriptions facilitate interpretation of archival records in addition to facilitating access. Understanding the lives and work of people and groups who created the archival collections is essential to understanding the archival records that are the byproducts of those lives and activities. In other words, EAC–CPF records attempt to place creators in context.

Data encoded with EAC–CPF is intended for use in federated database applications and collaborative research across a broad range of domains, including prosopographical research (i.e., research that identifies and relates a group of persons or characters within a particular historical or literary context) and genealogical studies. The designers intend that intellectual content of EAC–CPF records comply with the ISAAR(CPF). The EAC–CPF format complements other authority structures from a variety of domains. While coming from the archival community its structure does not provide any limitation as to the types of resources to which it may be connected.

The basic structure of EAC–CPF includes a root element, **<eac-cpf>**, with two sub-elements, **<control>** and **<cpfDescription>**. The control section includes information about the creation, maintenance, status, rules and authorities, and sources used to create the EAC–CPF record. The cpfDescription section contains sub-elements such as: *Identity*, *Description*, and *Relations*. Only the *Identity* element is required, and it is used to establish both authorized and variant names for the entity being described. It also includes a required *entity type* element to identify whether the record is a corporate body, a person, or a family. The *Description* element includes various controlled vocabulary terms and narrative descriptive elements, such as

- dates of existence
- functions
- legal status
- mandates
- occupations
- places
- organizational structure
- genealogy information
- biographical information
- historical information

The *Relations* element provides structures of relationships with other corporate bodies, persons, and families; with resources; and with descriptions of functions. Through this structure, it leverages linked data strategies and helps to facilitate connections between entities, resources, and the functions and activities that create or modify records.

Several archives-related projects have implemented the EAC–CPF standard. These range from single institution projects focused on specific topic areas to national and international efforts. One of the most developed in the use of EAC–CPF is Trove from the National Library of Australia.[45] The foundations of the Trove dataset were harvested from existing database projects developed by various communities. Harvesting and aggregating the data, Trove provides access not only to records about entities but also to various resource types. Archives Portal Europe is an online research tool developed collaboratively from 2009 to 2015 and funded by the European Commission.[46] It contains EAD and EAC–CPF records contributed by institutions throughout Europe. Social Networks and Archival Context (SNAC) is a collaborative project in the United States that aggregates information about entities in order to provide access to distributed archival resources.[47] SNAC began as a research project on the viability of harvesting existing data sources and storing that information in the EAC–CPF format. With projects such as these, the use of EAC–CPF continues to gain momentum.

STANDARDS FOR ART AND MUSEUMS

The art and museum community has recently begun to address the need for control of names and titles in descriptions of objects. The Getty Research Institute has created a set of controlled vocabulary tools[48] including one for terminology in the fields of art and architecture (*Art and Architecture Thesaurus*); one for artist names, including biographical information as well as variations in names (*Union List of Artist Names*); one for geographic names, including vernacular and historical names, coordinates, and place types (*Getty Thesaurus of Geographic Names*), and one for names of cultural objects (*Cultural Objects Name Authority*). The art and museum community has also created standards that provide for consistency in its metadata. The standards addressed in this section are:

- *Cataloging Cultural Objects* (CCO)
- *Categories for the Description of Works of Art* (CDWA)
- VRA (Visual Resources Association) Core

Cataloging Cultural Objects (CCO)

Cataloging Cultural Objects (CCO)[49] is discussed in Chapter 7 as being a content standard rather than a prescription of metadata elements. The authors of CCO believe that authority control is a critical aspect of effective description, especially in the online environment. They recommend using authority files for certain metadata elements. Part three of CCO, "Authorities," covers guidelines for the elements that the authors believe are the most important to be controlled: *personal and corporate names*, *geographic places*, *concepts*, and *subjects*. Each of these four kinds of entities is discussed separately, with sections under each group for introductory material, "Editorial Rules," and "Presentation of the Data." The first three are discussed here. Subject authority (i.e., use of AAT) is discussed in Chapter 10.

Personal and corporate names are discussed together and cover artists, architects, studios, architectural firms, patrons, repositories, and "others responsible for the design and production of cultural works."[50] The following elements are recommended for name authority records (those in boldface type are required):

- **Names** (preferred, alternates, and variants)
- Note
- **Display Biography**
- **Birth Date** (or Start Date for corporate bodies)
- **Death Date** (or End Date for corporate bodies)
- **Nationality** (or National Affiliation for corporate bodies)

- **Life Roles** (or Functions for corporate bodies)
- Gender (not applicable for corporate bodies)
- Date of Earliest Activity
- Date of Latest Activity
- Place/Location
- Related People and Corporate Bodies
- Relationship Type
- Events
- **Sources**
- Record Type (Person or Corporate Body)

Some of the elements are free-text fields, while others are suggested to have prescribed formatting (e.g., date fields should use the ISO [International Organization for Standardization] standard for dates), and controlled vocabulary is recommended for others. A portion of a CCO authority record is presented in Figure 8.6.

The geographic names authority section covers places that are both physical features and administrative geographic entities. Geographic place names are used primarily for the location of works but are also used in authority records for personal and corporate names. The following are recommended elements for geographic place authority records (required elements are in boldface type):

Names: **Meléndez, Luis** (preferred, inverted)

Luis Meléndez (preferred, natural order)
Meléndez, Luis Egidio (variant)
Meléndez de Rivera Durazo y Santo Padre, Luis Egidio (variant)
Rivera Durazo y Santo Padre, Luis Egidio Meléndez de (variant)
...

Display Biography: Spanish painter, 1716-1780
Nationality: Spanish
Birth Date: 1716
Death Date: 1780
Life Roles: painter
miniaturist
Gender: male
Place of Birth: Naples (Campania, Italy)
Place of Death: Madrid (Comunidad de Madrid, Spain)

Related People:
Relationship Type: child of
Meléndez, Francisco Antonio (Spanish painter, 1682-1752)
Relationship Type: student of
Meléndez, Francisco Antonio (Spanish painter, 1682-1752)
Relationship Type: nephew of
Menéndez, Miguel Jacinto (Spanish painter, 1679-1734)
Relationship Type: assistant of
Loo, Louis Michel van (French painter, 1707-1771)

Note: Luis Meléndez (1715–1780) is now recognized as the premier still-life painter in 18th-century Spain, indeed one of the greatest in all of Europe, though his reputation had long been eclipsed by the achievements of his Spanish contemporary, Francisco Goya. ...

Sources: Union List of Artist Names (1988-).

National Gallery of Art, "Luis Melendez: Master of the Spanish Still Life: May 17 – August 23, 2009." https://www.nga.gov/exhibitions/melendezinfo.shtm.

Figure 8.6. A Representation of a CCO Authority Record for a Personal Name.

- **Names** (preferred, alternates, and variants)
- **Broader Context**

- **Place Type**
- Coordinates
- Note
- Related Places
- Relationship Type
- Dates
- **Sources**

Some elements are controlled, some are free-text, and for *Coordinates* and *Dates*, especially, it is recommended that consistent formatting be used.

Concept authority includes terminology needed for describing works or images that is not included in the other authorities (i.e., not names of persons, organizations, geographic places, events, and not subjects). Concepts may describe the type of work (e.g., painting), its material (e.g., canvas), activities associated with a work (e.g., oil painting [technique]), its style (e.g., Art Nouveau), the role of the creator or other persons (e.g., painters [artists]), and other attributes or various abstract concepts (e.g., contrast). The following are recommended elements for concept authority records (required elements are in boldface type):

- **Terms** (preferred, alternates, and variants)
- Qualifier
- **Broader Context**
- **Note**
- Dates
- Related Concepts
- Relationship Type
- **Sources**

Concept authority records, in particular, are created in such a way that the terms in them can be displayed in a hierarchical arrangement.

CCO lists a number of sources for authority files for names, geographic places, and concepts. Among the most important in the world of art, architecture, and museums are the previously mentioned Getty Vocabularies:[51] *Union List of Artist Names* (ULAN), *Getty Thesaurus of Geographic Names* (TGN), *Art & Architecture Thesaurus* (AAT), and *Cultural Objects Name Authority* (CONA).

In addition to its emphasis on authority control for some fields, CCO puts a premium on identifying relationships between works.[52] CCO distinguishes between intrinsic and extrinsic relationships. An *intrinsic relationship* is one that is essential to identification of the work being cataloged. Whole–part relationships, for example, are intrinsic. Examples of whole–part relationships include architectural complexes (e.g., the statues atop a building described separately from the building), manuscripts (e.g., the illustrations described separately from the text), triptychs (e.g., separate description of one panel of an altarpiece of three panels), works in collections, and items in a series. An *extrinsic relationship* is one that is informative but not essential. Examples of extrinsic relationships include a work copied after another, a preparatory sketch or model, a work referenced within another work (e.g., a famous painting on the wall behind the main object in another painting), or two works that are intended to be seen together (e.g., separate portraits of a husband and wife). CCO emphasizes displaying relationships in a way that is clear to the user. Attention is given to the concept of database design and to the display of relationships when designing and constructing databases.[53]

Categories for the Description of Works of Art (CDWA)

Categories for the Description of Works of Art (CDWA)[54] is discussed in Chapter 7 as an element set for describing and accessing works of art, architecture, and other material culture. Fifteen of the 36 core categories of CDWA are listed in Chapter 7. The remaining 21 are listed here, divided into five groupings as organized by CDWA for authority data:

- **For Creator Identification Authority**
  - Name
  - Name Source
  - Display Biography
  - Birth Date
  - Death Date
  - Nationality/Culture/Race
  - Life Roles
- **For Place/Location Authority**
  - Place Name
  - Place Name Source
  - Place Type
  - Broader Context
- **For Related Textual References Authority**
  - Brief Citation
  - Full Citation
- **For Generic Concept Authority**
  - Term
  - Term Source
  - Broader Context
  - Scope Note
  - Note Source
- **For Subject Authority**
  - Subject Name
  - Source
  - Broader Context

These elements are much like the ones listed in CCO. Each of the five groupings contains a much larger list of elements that may be used to describe more specifically that entity type. For example, under *creators* one finds the following elements:

- 28.1. Person Authority Record Type
- 28.2. Person/Corporate Body Name
  - 28.2.1. Preference
  - 28.2.2. Name Type
  - 28.2.3. Name Qualifier
  - 28.2.4. Name Language
  - 28.2.5. Historical Flag
  - 28.2.6. Display Name Flag
  - 28.2.7. Other Name Flags
  - 28.2.8. Name Source
  - 28.2.9. Name Date

- 28.3. Display Biography
- 28.4. Birth Date
- 28.5. Death Date
- 28.6. Birth Place
- 28.7. Death Place
- 28.8. Person Nationality/Culture/Race
- 28.9. Gender
- 28.10. Life Roles
- 28.11. Person/Corporate Body Event
- 28.12. Related Person/Corporate Body
- 28.13. Person/Corporate Body Broader Context
- 28.14. Person/Corporate Body Label/Identification
- 28.15. Person/Corporate Body Descriptive Note
- 28.16. Remarks
- 28.17. Citations
- 28.18. Person Authority Record ID

Additional sub-elements appear under more than a few of those listed.

CDWA also recommends maintaining separate authority files for visual works, textual materials, persons/corporate bodies, locations/places, generic concepts, and subjects so that such information is recorded once and then linked to all appropriate work records.

VRA (Visual Resources Association) Core

The VRA Core[55] was discussed in Chapter 7 as a "data standard for the description of works of visual culture as well as the images that document them."[56] The 19 individual metadata elements of the VRA Core were listed in the previous chapter (in alphabetical order, except for the first element), along with suggestions as to which elements could take their values from an authority file or controlled vocabulary list.

VRA Core, like CCO, emphasizes relationships in its implementation. Its basic data model stresses that descriptions may focus on one of three primary divisions—*works*, *images* of those works, and *collections*. Thus, the three primary types of relationships between records in VRA Core 4.0 are work-to-work, image-to-work, and image- or work-to-collection relationships. The VRA Core introduction document explains that in order to give meaning to an image record, it is necessary to show its relationship to the work from which it was derived. The link is made via the *relation* element, and the nature of the relationship is defined by the *type* attribute. A work may have several images associated with it, or an image may represent more than one work; such multiple relationships should be identified. In VRA Core, *collection* is defined as an aggregate of works or images. Therefore, it may be necessary to show image-to-collection relationships or work-to-collection relationships. Typically VRA Core is not used to create authority data for agents. *Agent* is provided as an attribute for a work or an image. Sub-elements under agent, however, include *name*, *culture*, *dates*, *role*, and *attribution*.

STANDARDS AND PROJECTS IN ONLINE SETTINGS

During the last two decades several research projects have looked at ways to describe entities/agents/creators/parties in such a way as to avoid duplication of effort and to link identical and complementary identities. A research project that began in the late 1990s and ended in 2003 was called InterParty.[57] Even though it did not produce an active namespace, it had some interesting characteristics that are thought provoking for anyone contemplating future uses of authority files. The project defined *party* as any one of the disparate types of identities responsible for creation of intellectual property or content (e.g., authors, composers, performers, producers, directors, publishers, libraries). The project brought together people from the book industry, music recording industry, rights management, libraries, the technology community, and the identifier community. InterParty looked at the fact that databases containing metadata about people and organizations already exist and identify parties in their own context (e.g., publisher databases of authors, editors; music industry databases of composers, directors; library authority files of authors, editors, composers). InterParty suggested that such databases could be linked with a common functional goal of unique identification and disambiguation of parties. Each database would agree to make available common nonconfidential metadata. InterParty would add a new layer of links that could give messages such as "Person X in Namespace A is the same [or not the same] as Person Y in Namespace B." The owner of Namespace B would then need to verify such an assertion by agreeing or disputing. While this research effort did not yield a separate standard to reflect its findings, much of what was uncovered has found expression in standards and schemas developed in the past 10 years, particularly VIAF.

Another project, begun in 1998 and conducted by the Dublin Core community, recognized the need for *agent metadata* to complement and be linked to metadata for information resources. The term *agent* was used to refer to creators, contributors, and publishers. This terminology, in recent years, has continued to be used as a generic term that encompasses various roles associated with the creation and production of resources. For example, it is a term that is being used in the IFLA LRM to act as the class of entities containing persons and collective agents (e.g., families and corporate bodies). The DCMI Agents Working Group[58] was working toward developing functional requirements for describing agents. They wished to identify existing conventions for agent description, one of which was the InterParty project described above. They intended to develop a recommendation for an agent element set, but the work has never come to fruition. The DCMI Agents Working Group was deactivated in 2009.

The *International Standard Name Identifier* (ISNI)[59] initiative, begun in 2011, is another effort to create persistent identifiers for persons and corporate bodies and facilitate identity management and disambiguation. ISNI is a standard, adopted by ISO, for the unique identification of public identities in all fields of creative activity. ISNI has three related objectives. It aims to create identifiers that

- are relevant across domains,
- serve all the needs of the various information industry stakeholders, and
- support a high level of interoperability.

ISNI is currently supported through a membership model and is governed by the ISNI International Agency.

ORCID (originally standing for Open Researcher and Contributor ID) was organized in 2010. It is a separate initiative to "connect research to researchers."[60] It is a nonprofit organization specifically targeting individuals engaged in scholarly communication, and it provides persistent unique identifiers for individual scholars to deal with issues such as name ambiguity associated with scholarly publications. ORCID is subscription based and contains information not only about scholars but also their affiliations and relationships to scholarly products.[61]

> ORCID is committed to being interoperable with other identifier schemes, including ISNI. To this end, ORCID and ISNI are coordinating their efforts where they overlap in the research and scholarship communities. ORCID identifiers utilize a format compliant with the ISNI ISO standard. ISNI has reserved a block of identifiers for use by ORCID, so there will be no overlaps in assignments.[62]

The ORCID and ISNI organizations are looking for additional opportunities for collaboration.

CONCLUSION

This chapter has discussed the creation of name and title access points for resource descriptions. The massive amounts of data dealt with in libraries, archives, and museums for art, architecture, and cultural objects have led to the development of the concept of authority control. Searchers for information need to be able to find works related to specific persons, corporate bodies, places, or other works. As the number of names and titles increases, the ability to find specific names and titles and sort them out from similar and identical ones becomes very difficult without authority control. The building of the Semantic Web has the promise of enhancing the advantages of authority work by linking authority files with a diverse range of other resources and data sources, such as biographical dictionaries, genealogical databases, telephone directories, and official websites for named entities, as well as with the information resources created by those entities.

Some Important Terms in This Chapter

Access Point	Expression	Nomen
Added Entry	Extrinsic Relationship	Party
Agent	Four-Fold Path	Preferred Name
Asynchronous Responsibility	Full Entry	Preferred Title for Work
Authorized Access Point	Find	Primary Access Point
Authority Control	Heading	Relationship Designator
Authority File	Identifier	Res
Collective Agent	Identify	Rule of Three
Contextualize	Intrinsic Relationship	Synchronous Responsibility
Contributor	Item	Thema
Controlled Access Point	Justify	Title Proper
Controlled Title	Keyword	Uncontrolled Title
Core Element	Main Entry	Uniform Title
Creator	Manifestation	Variant Access Point
Data Value Standards	Name	Variant Name

Most information resources have a subject or multiple subjects. It is often the subject content that is sought by users of retrieval tools. The next three chapters address many ways of providing access to subject content.

Some Important Acronyms in This Chapter

AACR:	*Anglo-American Cataloging Rules*
AACR2:	*Anglo-American Cataloguing Rules, Second edition*
AAP:	Authorized Access Point
AAT:	*Art & Architecture Thesaurus*
CCO:	*Cataloging Cultural Objects*
CDWA:	*Categories for the Description of Works of Art*
CONA:	*Cultural Objects Name Authority*
DACS:	*Describing Archives: A Content Standard*
EAC–CPF:	Encoded Archival Context–Corporate Bodies, Persons, and Families
FRAD:	*Functional Requirements for Authority Data*
FRBR:	*Functional Requirements for Bibliographic Records*
FRSAD:	*Functional Requirements for Subject Authority Data*
GARR:	*Guidelines for Authority Records and References*
ICA:	International Council on Archives
ICP:	*Statement of International Cataloguing Principles*
IFLA:	International Federation of Library Associations and Institutions
IRI:	Internationalized Resource Identifier
ISAAR(CPF):	*International Standard Archival Authority Record for Corporate Bodies, Persons, and Families*
ISAD(G):	*General International Standard Archival Description*
ISNI:	International Standard Name Identifier
ISO:	International Organization for Standardization
LCC:	*Library of Congress Classification*
LCDGT:	*Library of Congress Demographic Group Terms*
LCGFT:	*Library of Congress Genre/Form Terms for Library and Archival Materials*
LCSH:	*Library of Congress Subject Headings*
LRM:	*IFLA Library Reference Model*
MADS:	Metadata Authority Descriptions Schema
MARC:	MAchine-Readable Cataloging
ORCID:	Open Researcher and Contributor ID
RDA:	*Resource Description & Access*
TGN:	*Getty Thesaurus of Geographic Names*
ULAN:	*Union List of Artist Names*
URI:	Uniform Resource Identifier
VAP:	Variant Access Point
VRA:	Visual Resources Association
WEMI:	Work, Expression, Manifestation, Item

NOTES

All URLs accessed June 2017.

1. Barbara B. Tillett, "A Taxonomy of Bibliographic Relationships," *Library Resources & Technical Services* 35, no. 2 (April 1991): 156.

2. For more in-depth treatments of bibliographic relationships, see: Elaine Svenonius, *The Intellectual Foundation of Information Organization* (Cambridge, MA: MIT Press, 2001), 98–106; and Richard P. Smiraglia, *The Nature of "A Work": Implications for the Organization of Knowledge* (Lanham, MD: Scarecrow, 2001), 35–52.

3. Smiraglia, *Nature of "A Work,"* 42.

4. Svenonius, *Intellectual Foundation of Information Organization,* 95–97.

5. Martha M. Yee, "What Is a Work?" [a four-part series published in *Cataloging & Classification Quarterly* 19, nos. 1–2 (1994), and 20, nos. 1–2 (1995)]; Smiraglia, *Nature of "A Work"*; International Federation of Library Associations and Institutions, IFLA Study Group, *Functional Requirements for Bibliographic Records: Final Report* (Munich: Saur, 1998), http://www.ifla.org/publications/functional-requirements-for-bibliographic-records.

6. Amazon.com: Online Shopping for Electronics, Apparel, Computers, Books, DVDs, & More, www.amazon.com.

7. Library of Congress Online Catalog, https://catalog.loc.gov/.

8. Program for Cooperative Cataloging, "NACO: Name Authority Co-operative Program," http://www.loc.gov/aba/pcc/naco/index.html.

9. Library of Congress, Linked Data Service: Authorities and Vocabularies, http://id.loc.gov/.

10. OCLC, "VIAF: The Virtual International Authority File," http://www.oclc.org/research/activities/viaf.html.

11. Barbara Tillett, "Authority Control: State of the Art and New Perspectives," *Cataloging & Classification Quarterly* 38, no. 3/4 (2004): 23–42; also published in *Authority Control in Organizing and Accessing Information: Definition and International Experience,* eds. Arlene G. Taylor and Barbara B. Tillett (New York: Haworth Information Press, 2004), 23–42.

12. OCLC, "VIAF: The Virtual International Authority File," http://www.oclc.org/research/activities/viaf.html.

13. OCLC, "VIAF (Virtual International Authority File)," http://www.oclc.org/developer/develop/web-services/viaf.en.html.

14. IFLA, *Guidelines for Authority Records and References,* 2nd ed. (Munich: K. G. Saur, 2001), http://www.ifla.org/publications/ifla-series-on-bibliographic-control-23.

15. International Federation of Library Associations and Institutions (IFLA) Working Group on Functional Requirements and Numbering of Authority Records (FRANAR), *Functional Requirements for Authority Data: A Conceptual Model*, approved by the Standing Committees of the IFLA Cataloguing Section and IFLA Classification and Indexing Section, 2009, amended and corrected through July 2013, http://www.ifla.org/publications/functional-requirements-for-authority-data.

16. International Federation of Library Associations and Institutions, Working Group on the Functional Requirements for Subject Authority Records (FRSAR), *Functional Requirements for Subject Authority Data: A Conceptual Model*, approved by the Standing Committee of the IFLA Section on Classification and Indexing, 2010, http://www.ifla.org/node/5849.

17. Pat Riva, Patrick Le Boeuf, and Maja Žumer, "IFLA Library Reference Model: A Conceptual Model for Bibliographic Information," August 2017 (The Hague: IFLA, 2017), https://www.ifla.org/publications/node/11412.

18. Open Archives Initiative, "Open Archives Initiative Protocol for Metadata Harvesting," http://www.openarchives.org/pmh/.

19. IFLA, FRAD, http://www.ifla.org/publications/functional-requirements-for-authority-data.

20. IFLA, FRSAD, http://www.ifla.org/node/5849.

21. For more about FRBR see: Daniel N. Joudrey, Arlene G. Taylor, and David P. Miller, *Introduction to Cataloging and Classification*, 11th ed. (Santa Barbara, CA: Libraries Unlimited, 2015), Chapter 3; Arlene G. Taylor, "An Introduction to Functional Requirements for Bibliographic Records (FRBR)," in *Understanding FRBR: What It Is and How It Will Affect Our Retrieval Tools,* ed. Arlene G. Taylor (Westport, CT: Libraries Unlimited, 2007), 1–19; and Robert L. Maxwell, *FRBR: A Guide for the Perplexed* (Chicago: American Library Association, 2008).

22. IFLA, FRAD, 30–49, http://www.ifla.org/publications/functional-requirements-for-authority-data.

23. Ibid., 31–32.

24. Riva, Le Boeuf, and Žumer, "IFLA Library Reference Model," https://www.ifla.org/publications/node/11412, 9.

25. Ibid.

26. International Federation of Library Associations and Institutions (IFLA), Meeting of Experts on an International Cataloguing Code, *Statement of International Cataloguing Principles* (ICP), 2016 edition (The Hague: IFLA, 2016), http://www.ifla.org/publications/node/11015.

27. IFLA, "IME-ICC: IFLA Meetings of Experts on an International Cataloguing Code," http://www.ifla.org/node/576.

28. IFLA, ICP, 4.

29. Ibid., 5.

30. Ibid., 8.

31. Ibid., 8–10.

32. Joint Steering Committee for Development of RDA, "Historic Documents: Outcomes of the Meeting of the Joint Steering Committee Held in Washington, DC, 16–20 October 2006," http://www.rda-jsc.org/archivedsite/0610out.html.

33. "Works" is what is referred to as a *conventional collective title*. It is used for the complete works of an author. Other conventional collective titles include terms such as *Plays*, *Poems*, *Correspondence*, *Speeches*, *Selections*, and so on.

34. International Conference on Cataloguing Principles (Paris: 1961), *Report* (London: International Federation of Library Associations, 1963), 91–96. Also available in: *Library Resources & Technical Services* 6 (1962): 162–67; and *Statement of Principles Adopted at the International Conference on Cataloguing Principles, Paris, October, 1961,* annotated ed., with commentary and examples by Eva Verona (London: IFLA Committee on Cataloguing, 1971); and A. H. Chaplin and Dorothy Anderson, eds., *Report/International Conference on Cataloguing Principles, Paris, 9th–18th October 1961* (London: IFLA International Office for UBC, 1981).

35. Doris Hargrett Clack, ed., *The Making of a Code: The Issues Underlying AACR2* (Chicago: American Library Association, 1980).

36. Michael Gorman, "AACR2: Main Themes," in *The Making of a Code: The Issues Underlying AACR2*, ed. Doris Hargrett Clack (Chicago: American Library Association, 1980), 46.

37. Richard P. Smiraglia, "Authority Control and the Extent of Derivative Bibliographic Relationships" (PhD diss., University of Chicago, 1992).

38. Arlene G. Taylor, *Introduction to Cataloging and Classification*, 10th ed. with the assistance of David P. Miller (Westport, CT: Libraries Unlimited, 2006).

39. Library of Congress, "MADS: Metadata Authority Description Schema," http://www.loc.gov/standards/mads/.

40. International Council on Archives, *International Standard Archival Authority Record for Corporate Bodies, Persons, and Families* (ISAAR(CPF)), 2nd ed. (Paris: International Council on Archives, 2004).

41. *Describing Archives: A Content Standard* (DACS) (Chicago: Society of American Archivists, 2004), 87, also available at: http://www2.archivists.org/groups/technical-subcommittee-on-describing-archives-a-content-standard-dacs/dacs.

42. Ibid., 92–93.

43. Society of American Archivists, "Encoded Archival Context: Corporate Bodies, Persons, and Families (EAC–CPF)," http://eac.staatsbibliothek-berlin.de/.

44. Daniel V. Pitti, "Creator Description: Encoded Archival Context," *Cataloging & Classification Quarterly* 39, no. 1/2 (2004): 201–26; also published in *Authority Control in Organizing and Accessing Information: Definition and International Experience,* eds. Arlene G. Taylor and Barbara B. Tillett (New York: Haworth Information, 2004), 201–26.

45. National Library of Australia, "Trove," http://trove.nla.gov.au/.

46. Archives Portal Europe Foundation, "Welcome to the Archives Portal Europe Foundation," http://www.archivesportaleuropefoundation.eu/.

47. Social Networks and Archival Context (SNAC), http://socialarchive.iath.virginia.edu/collaborators.html.

48. The Getty, "Getty Vocabularies," http://www.getty.edu/research/tools/vocabularies/.

49. Murtha Baca, Patricia Harpring, Elisa Lanzi, Linda McRae, and Ann Whiteside, on behalf of the Visual Resources Association, *Cataloging Cultural Objects: A Guide to Describing Cultural Works and Their Images* (Chicago: American Library Association, 2006), also available at: http://cco.vrafoundation.org/ [cited hereinafter as CCO].

50. CCO, 279.

51. The Getty, "Getty Vocabularies," http://www.getty.edu/research/tools/vocabularies/.

52. CCO, 13–19.

53. CCO, 20–27.

54. The Getty, "Categories for the Description of Works of Art," eds. Murtha Baca and Patricia Harpring; rev. March 2014 by Patricia Harpring, http://www.getty.edu/research/publications/electronic_publications/cdwa/.

55. Library of Congress, "VRA Core: A Data Standard for the Description of Images and Works of Art and Culture," https://www.loc.gov/standards/vracore/.

56. Library of Congress, "An Introduction to VRA Core," https://www.loc.gov/standards/vracore/VRA_Core4_Intro.pdf.

57. Andrew MacEwan, "Project InterParty: From Library Authority Files to E-Commerce," *Cataloging & Classification Quarterly* 39, no. 1/2 (2004): 429–42; also published in *Authority Control in Organizing and Accessing Information: Definition and International Experience,* eds. Arlene G. Taylor and Barbara B. Tillett (New York: Haworth Information, 2004), 429–42.

58. "DCMI Agents Working Group," Dublin Core Metadata Initiative, http://www.dublincore.org/groups/agents/.

59. International Standard Name Identifier, http://www.isni.org/.

60. ORCID, https://orcid.org/about/what-is-orcid/mission.

61. Ibid.

62. ORCID, "What Is the Relationship between ISNI and ORCID?" https://support.orcid.org/knowledgebase/articles/115265-what-is-the-relationship-between-isni-and-orcid.

SUGGESTED READINGS

Baca, Murtha, Patricia Harpring, Elisa Lanzi, Linda McRae, and Ann Whiteside. *Cataloging Cultural Objects: A Guide to Describing Cultural Works and Their Images.* Chicago: American Library Association, 2006. Also available at: http://cco.vrafoundation.org/.

Chan, Lois Mai, and Athena Salaba. *Cataloging and Classification: An Introduction.* 4th ed. Lanham, MD: Rowman & Littlefield, 2016. Chapters 7–9.

Describing Archives: A Content Standard (DACS). 2nd ed. Chicago: Society of American Archivists, 2013. http://www2.archivists.org/groups/technical-subcommittee-on-describing-archives-a-content-standard-dacs/dacs.

The Getty. "Getty Vocabularies." http://www.getty.edu/research/tools/vocabularies/index.html.

Harpring, Patricia. *Categories for the Description of Works of Art*, edited by Murtha Baca and Patricia Harpring. Revised March 2014 by Patricia Harpring. Los Angeles: The Getty, 2014. http://www.getty.edu/research/publications/electronic_publications/cdwa/.

International Federation of Library Associations and Institutions, IFLA Study Group. *Functional Requirements for Bibliographic Records: Final Report.* Munich: Saur, 1998. https://www.ifla.org/publications/functional-requirements-for-bibliographic-records.

International Federation of Library Associations and Institutions, Meetings of Experts on an International Cataloguing Code. *Statement of International Cataloguing Principles*, 2016 edition. The Hague: IFLA, 2016. http://www.ifla.org/files/assets/cataloguing/icp/icp_2016-en.pdf.

International Federation of Library Associations and Institutions, Working Group on Functional Requirements and Numbering of Authority Records (FRANAR). *Functional Requirements for Authority Data: A Conceptual Model*, approved by the Standing Committees of the IFLA Cataloguing Section and IFLA Classification and Indexing Section, 2009, amended and corrected through July 2013. http://www.ifla.org/publications/functional-requirements-for-authority-data.

Intner, Sheila S., and Jean Weihs. *Standard Cataloging for School and Public Libraries.* 5th ed. Santa Barbara, CA: Libraries Unlimited, 2015. Chapter 5: "Name and Title Access Points."

Joudrey, Daniel N., Arlene G. Taylor, and David P. Miller. *Introduction to Cataloging and Classification*. 11th ed. Santa Barbara, CA: Libraries Unlimited, 2015. Chapter 6: "Works and Expressions," Chapter 7: "Persons, Families, Places, and Corporate Bodies," Chapter 8: "Relationships and the Use of Access Points," and Chapter 10: "Authority Control."

Maxwell, Robert L. *Maxwell's Handbook for RDA, Resource Description & Access: Explaining and Illustrating RDA: Resource Description and Access Using MARC21*. Chicago: American Library Association, 2013.

Moulaison, Heather Lea, and Raegan Wiechert. *Crash Course in Basic Cataloging with RDA.* Santa Barbara, CA: Libraries Unlimited, 2015. Chapter 7: "Access Points and Controlled Vocabulary."

Patton, Glenn E. "An Introduction to Functional Requirements for Authority Data (FRAD)" and "Understanding the Relationship between FRBR and FRAD." In *Understanding FRBR: What It Is and How It Will Affect Our Retrieval Tools,* edited by Arlene G. Taylor, 21–33. Westport, CT: Libraries Unlimited, 2007.

Pitti, Daniel V. "Creator Description: Encoded Archival Context." *Cataloging & Classification Quarterly* 39, no. 1/2 (2004): 201–26; also published in *Authority Control in Organizing and Accessing Information: Definition and International Experience,* edited by Arlene G. Taylor and Barbara B. Tillett, 201–26. New York: Haworth Information, 2004.

Smiraglia, Richard P. "Bibliographic Families and Superworks." In *Understanding FRBR: What It Is and How It Will Affect Our Retrieval Tools,* edited by Arlene G. Taylor, 73–86. Westport, CT: Libraries Unlimited, 2007.

Smiraglia, Richard P. *The Nature of "A Work": Implications for the Organization of Knowledge.* Lanham, MD: Scarecrow, 2001.

Svenonius, Elaine. *The Intellectual Foundation of Information Organization.* Cambridge, MA: MIT Press, 2000. Chapter 3: "Bibliographic Entities" and Chapter 6: "Work Languages."

Taylor, Arlene G. *Introduction to Cataloging and Classification.* 10th ed. With the assistance of David P. Miller. Westport, CT: Libraries Unlimited, 2006. Chapter 6: "Choice of Access Points," Chapter 7:

"Form of Headings for Names and Titles," and Chapter 8: "Authority Control."

Taylor, Arlene G. "An Introduction to Functional Requirements for Bibliographic Records (FRBR)." In *Understanding FRBR: What It Is and How It Will Affect Our Retrieval Tools,* edited by Arlene G. Taylor, 1–19. Westport, CT: Libraries Unlimited, 2007. [Other articles in this volume include discussions of FRAD, history of FRBR, research on FRBR, bibliographic families, and discussions of FRBR in relation to RDA, archival materials, art/architecture/material culture, cartographic materials, music, and serials.]

Tillett, Barbara B. "Authority Control: State of the Art and New Perspectives." *Cataloging & Classification Quarterly* 38, no. 3/4 (2004): 23–42; also published in *Authority Control in Organizing and Accessing Information: Definition and International Experience,* edited by Arlene G. Taylor and Barbara B. Tillett, 23–42. New York: Haworth Information, 2004.

Tillett, Barbara B. "Bibliographic Relationships." In *Relationships in the Organization of Knowledge,* edited by Carole A. Bean and Rebecca Green, 9–35. Dordrecht, The Netherlands: Kluwer Academic, 2001.

Wilson, Patrick. "The Catalog as Access Mechanism: Background and Concepts." In *Foundations of Cataloging: A Sourcebook,* edited by Michael Carpenter and Elaine Svenonius, 253–68. Littleton, CO: Libraries Unlimited, 1985.

CHAPTER 9

SUBJECT ANALYSIS

Historically, subject access has been one of the most challenging aspects of organizing information. *Subject analysis*—identifying and describing what a resource is and what it is about—can be difficult and time-consuming work even with the most traditional of information resources. With nontextual, imaginative, or complex scholarly materials, the process can be even more demanding. Despite the difficulties and costs associated with subject analysis, information professionals continue to see the value inherent in identifying precisely a resource's form and subject matter (often referred to as *aboutness* in library and information science literature), and then carefully choosing the most suitable terms and symbols from controlled lists to represent the resource's aboutness in its metadata.

In recent years, with advances in search engine technology and the high costs of original cataloging, the necessity for subject analysis has been questioned. Some have suggested that information resources no longer need to be analyzed because when users are searching for information, computers (through automatic indexing) can identify *some* documents relevant to users' needs (although not all), and therefore the time and money spent on humans performing subject analysis could be diverted to other activities such as digitization projects. Others have suggested that computers can analyze documents and assign classification numbers and/or descriptors from a list of controlled vocabulary terms. Despite improvements in search engines, the development of *microdata* (see Chapter 1), and the introduction of user tagging and other activities meant to improve online retrieval, many library and information science (LIS) professionals are reluctant to turn over subject analysis activities to machines. There is still a preference in our discipline for reflective subject analysis over algorithmic approaches to subject access.

Why is there still such reluctance? Despite some advances, technology, thus far, has not proved itself up to the tasks of subject analysis. Machines are not yet adept at identifying the aboutness of information resources and they still cannot, with any satisfactory degree of accuracy, assign concepts from *subject languages* (i.e., controlled vocabularies and classification schemes). While a computer can determine what words are used in a document and the frequency of those words, at this time it cannot understand the multifaceted or nuanced concepts represented by those words. Even the most sophisticated algorithms cannot replace the human mind for its efficiency and efficacy in understanding the deeper meanings of texts and being able to represent those meanings through the use of subject languages. Machines simply do not provide thoughtful analysis, they do not have insights into what content is most significant or meaningful, and they cannot prioritize subject concepts, other than ranking them based on word-frequency (i.e., a process that is computational, not conceptual). In 2008 the Working Group on the Future of Bibliographic Control, a group organized by the Library of Congress (LC) to examine the role of cataloging in the twenty-first century, reaffirmed the importance and need for this human-centric task: "Subject analysis—including analyzing content and creating and applying subject headings and classification numbers—is a core function of cataloging; although expensive, it is nonetheless critical."[1]

In libraries, archives, museums, and indexing environments, it is still necessary for humans to determine what an information resource or a collection is about and which concepts are to be represented in its metadata. This approach, however, also has its share of problems. Although LIS researchers have been prolific in exploring the usefulness and limitations of subject languages, they have produced very little on how to actually determine aboutness. Ironically, the task most dependent on humans is the one least explored and understood! Consequently, many information professionals have only had instruction in the application of various controlled vocabularies and not in the process of determining aboutness. This can lead to difficulties, especially when the topics of the resource are not self-evident or the resource has several meanings or purposes. Those who lack an understanding of the multiple components and approaches to determining aboutness may not identify key concepts and consequently may omit useful controlled vocabulary terms when describing a resource. Over the years, a number of writers have pointed out that aboutness is more than just the words found in the resource; determining aboutness may also involve determining the creator's purpose, thinking about who may use the resource and for what purposes, considering what questions the document may answer, and various other aspects. F. W. Lancaster, who was a leading authority on indexing and abstracting, went so far as to say that the same resource could be described in various ways in different institutions, and that it *should* be, if the intended users would be interested in the resource for different reasons.[2] Due to concerns related to costs and interoperability, most information institutions focus their subject analysis activities on the most visible and universal aspects of a resource's aboutness, rather than attempting to describe all possible uses of the resource or all possible interpretations or meanings.

Several questions are addressed in this chapter. They include: What is subject analysis? What are some of the difficulties associated with the subject analysis process? How is the process performed? What bibliographic features are useful in the determination of aboutness? How is subject content determined for nontextual materials?

WHAT IS SUBJECT ANALYSIS?

Subject analysis is the part of the metadata creation process that identifies and articulates the subject matter and genre/form properties of the information resource being described. The process includes three important steps:

(1) examining a resource to determine what it is about,

(2) describing the aboutness in a written statement, and

(3) using that statement to assign controlled vocabulary terms and/or classification notations.

The first step, the *conceptual analysis,* is a comprehensive examination of the physical properties and the intellectual or creative contents of an information resource. This is done to understand what the resource *is* (i.e., its genre or form)[3] and what the resource *is about* (i.e., its subject matter). When analyzing a book, for example, one must review the myriad bibliographic features found in the resource (e.g., title page, table of contents, introduction, various chapters, illustrations) in order to determine the resource's form as well as its subject matter. Throughout the analysis, notes are taken so that details are not lost. If we were performing a conceptual analysis of *Introduction to Cataloging and Classification*, we might start with the title page (Figure 9.1) and the table of contents (Figure 9.2) before moving on to more content-heavy features of the resource (such as the preface and the actual text).

INTRODUCTION TO CATALOGING AND CLASSIFICATION

Eleventh Edition

Daniel N. Joudrey, Arlene G. Taylor, and David P. Miller

Library and Information Science Text Series

LIBRARIES UNLIMITED™

An Imprint of ABC-CLIO, LLC

Santa Barbara, California • Denver, Colorado

Figure 9.1. Title Page from Daniel N. Joudrey, Arlene G. Taylor, and David P. Miller, *Introduction to Cataloging and Classification*, 11th edition (Santa Barbara, CA: Libraries Unlimited, 2015).

Table of Contents

Figure 9.2. An Overview of the Table of Contents from *Introduction to Cataloging and Classification*.

In the second step, after the analysis is complete, an *aboutness statement* is written based on the notes that were taken. This may be a single sentence or a short paragraph describing and summarizing one's understanding of the aboutness, the genre/form, and the relationships among the important subject concepts.

The aboutness statement is used to identify the terms or concepts to be searched in the controlled vocabulary. The statement can also help the information professional to construct a rudimentary hierarchy to get a better understanding of how the concepts fit together and how they might fit into a scheme of classification. The following is an example of a brief aboutness statement (with significant concepts highlighted):

> This resource is a **textbook** about **library cataloging**. It addresses **descriptive cataloging** activities, with a particular focus on using the content standard ***RDA: Resource Description & Access***. There is also a significant section on **subject cataloging**, addressing both **subject headings** and **classification**.

The third and final step entails translating the aboutness statement into the specific symbols and/or terminology found in the subject languages employed. For example, in the third step, one would assign appropriate, authorized terms found in *Library of Congress Subject Headings* (LCSH), *Sears List of Subject Headings, Art & Architecture Thesaurus* (AAT), *Library of Congress Genre/Form Terms* (LCGFT), or another such list, and/or classification notation(s) from *Dewey Decimal Classification* (DDC), *Library of Congress Classification* (LCC), *Universal Decimal Classification* (UDC), or another scheme. The subject-related metadata for *Introduction to Cataloging and Classification* might include:

LCSH:	Descriptive cataloging.
	Subject cataloging.
	Classification-Books.
	Resource description & access.
LCGFT:	Textbooks
DDC:	025.3
LCC:	Z693

Throughout this process, the goals are not only to identify the intellectual and creative contents of information resources but also to ensure that individual resources are carefully and purposefully positioned within a collection. In other words, subject analysis is performed to provide users with subject access to information, to collocate resources of a like nature, and to provide a logical location for similar tangible resources on the shelves. Subject analysis helps to alleviate retrieval problems associated with keywords and natural language through the predictable use of controlled terminology and symbols. Ultimately, the goal is to save the users' time by providing an intellectual framework for finding similar resources together. If we once again look at Cutter's objectives (or objects) of the catalog (illustrated in Figure 3.1 on page 103), we see that he considered subject and genre/form access to be important functions of a catalog. Not only did he want users to be able to find a known work on a certain subject, but he also wanted users to be able to find all the works that the library could offer on a particular subject or in a particular genre or form. Subject analysis makes this possible.

CHALLENGES IN SUBJECT ANALYSIS

Determining what an information resource is about can be difficult, and not everyone agrees on how it should be done or even where the difficulties lie. Patrick Wilson, a philosopher and LIS scholar, has discussed many of the challenges in the subject analysis process. He believed that one difficulty was imposed upon us by Cutter's second object of the catalog, which states that a catalog should show what a library has on a given subject. Wilson suggested that it is problematic to take Cutter's statement to mean that there is an obvious subject in every information resource and that we should be able to identify it as *the* subject of the work.[4]

Although some resources may have a single, easily determined subject, others may not be quite so clear; they may have several multifaceted themes, with complex relationships among the subtopics. The aboutness of a college textbook titled *An Introduction to Sociology* is fairly straightforward to analyze and describe, but another text such as *An Encyclopedia of Nineteenth Century Sociology* is a bit more complex, and a third text, *Comparing Methods in Sociological Research in Panama, Peru, and Spain,* is even more so. All three are about sociology *per se,* but the second text is more specifically a resource that is in a particular bibliographic form (an encyclopedia) about sociology in the nineteenth century. Because the content is focused on the origins and early development of sociology, it is obviously written from a historical perspective, while not being *about* history. This distinction has a certain subtlety that is learned through education in our present-day Western tradition; in other places and other times, history would have been considered the subject of anything historical regardless of the specific topic. The third resource is a more complex subject that involves not only the discipline of sociology but also a specific aspect of it, that is, *research.* The information resource, however, is not just about sociological research; it compares the methods used in sociological research in three different locations. This resource involves a discipline, a subtopic, several geographic locations, and a comparative relationship (and possibly a chronological element and other content characteristics that are not evident from the title). Although this resource is more complex than the first (*An Introduction to Sociology*), it is still relatively simple to analyze.

With the burgeoning relationships among various fields, topics, and ideas in this increasingly interdisciplinary world, the result can be some very challenging materials to analyze. Consequently, this increased level of complexity may affect one's ability to understand certain resources and to articulate what they are about. For example, a highly technical dissertation may be more difficult to analyze than a work of popular science, and a complex literary analysis that demonstrates the confluence of social, religious, and economic factors in altering the portrayal of wealthy families in eighteenth-century English fiction is considerably more difficult to analyze than a book introducing a new diet regimen.

Cultural Differences

There are numerous factors that influence the subject analysis process. Some are related to the nature of the resource being analyzed, some to the persons who perform the analysis, and some to the subject languages we use for translation. An understanding of the place of one's culture as well as one's education in determining subject matter is important. Some researchers, such as social psychologist Richard E. Nisbett, believe that even our most basic cognitive processes are culturally influenced.[5] Nisbett's research indicates that persons from different parts of the world are likely to interpret concepts or situations differently based on historical cultural differences. Although his research has focused on differences between Eastern and Western societies, his theories could apply to other geographic and cultural divisions. George Lakoff, a cognitive linguist, has written about the research of anthropologists Brent Berlin and Paul Kay on the understanding of color depending upon one's language. Berlin and Kay found that there are 11 basic color categories in English, but in some other languages there are fewer categories. In languages that have only two basic color terms, the terms are the equivalent of *black* and *white,* or *cool* and *warm.*[6] In other words, their research shows that something as basic as identifying colors can be influenced by culture and language.

When persons doing subject analysis have grown up in different cultures with different languages, they might not comprehend the world around them in the same way, which influences their perceptions of aboutness. LIS scholar D. W. Langridge provides another example when he comments upon the unconscious effect that must occur in the mind of a person accustomed to the former arrangement of the library in the People's University of China (now Renmin University of China), where all knowledge was divided into three groups: theory of knowledge, knowledge of the class struggle, and knowledge of the productive struggle.[7] Although differences among Western cultures are perhaps not quite so dissimilar as those between Western and non-Western cultures, we should remember that we may see things differently than the person next to us, depending upon such things as education, language, and cultural background. For example, one indexer might refer to a group of persons as *revolutionaries*, while another might refer to that same group as *terrorists*, and yet another might refer to them as *freedom fighters*. This illustrates how our subject languages may affect a user's understanding of the world and its information.

Our subject access tools, like other social constructs, are not without biases. Despite the best intentions among those overseeing and maintaining subject languages, general classification schemes and controlled vocabularies often reflect only (or primarily) the points of view of the dominant culture. An American classification scheme, for example, is likely to dedicate more space to Christianity than it would to Jainism. This focus on the dominant culture may create difficulties in accurately describing resources created by, for, and about people who do not identify as part of that culture (which many would assert is the culture of white, Christian, middle-class, educated, English-speaking, married, straight, cisgendered men). When information professionals cannot easily represent additional points of view in our descriptions due to inherent biases in the subject languages employed, it can have a negative effect on access in our catalogs, indexes, finding aids, and other retrieval tools. Accurate and respectful representation of different groups and cultures has been, and continues to be, a significant challenge in subject analysis.

Consistency

Another challenge associated with the subject analysis process is consistency. Evidence of the difficulty in consistently determining and articulating aboutness is found in a number of studies in which people have been asked to list terminology that they would use to search for specific resources. For example, in a 1954 study by Oliver Lilley, 340 students looked at six books and suggested an average of 62 different terms that could be used to search for each book.[8] In a 1992 study, Lourdes Collantes found "an average of 25.6 [topical] names per object or concept."[9] Please note: *this is not a failure of controlled vocabulary.* It is a failure of individuals to come up with the same natural language terms to describe the resource or to determine the same aboutness from a document. There is evidence that information professionals using the same controlled vocabulary and the same rules for applying it will produce consistent subject headings, as long as they have the same understanding of the aboutness to draw upon.[10]

Langridge believes that inconsistencies in subject analysis are the result of confusing aboutness with purpose, that is, mixing up the questions, "*What is it about?*" and "*What is it for?*"[11] Langridge thinks that if those questions remain clear and separate, then determining aboutness should be fairly easy. Even if one does not agree with his assessment that determining aboutness can be straightforward, his question—*What is it for?*—provides a valuable insight into the conceptual analysis process. For example, Richard W. Unger's *The Art of Medieval Technology: Images of Noah the Shipbuilder* appears at first glance to be about Noah and the ark.[12] However, upon examining the author's purpose, the table of contents, and the captions with the illustrations, one learns that the work is really about changes in the techniques used in shipbuilding during the Middle Ages, as documented through medieval religious art, noting that artists' depictions of Noah building the ark changed as the technology and shipbuilding techniques advanced through the centuries.

Another example is found in the sociology examples mentioned earlier. In our culture, we consider the subject of *An Encyclopedia of Nineteenth Century Sociology* and *An Introduction to Sociology* to be the same: *sociology*. One resource, however, is written from the perspective of history. The two resources convey different types of information, and their purposes, perspectives, and forms affect our understanding of the aboutness. A user looking for one of these treatments would likely not be satisfied with the other. Addressing the resource's form, the creator's point of view, or the question "*What is it for?*" helps us to separate and distinguish among different treatments of the same subject. Addressing these types of issues, however, does not solve all of the challenges in the subject analysis process.

Nontextual Information

Determining the topics of nontextual information resources is even less clear-cut than the process for textual ones. For visual resources, several levels of conceptual analysis are possible. In 1939 art historian Erwin Panofsky identified three levels of meaning in works of art.[13]

(1) **The primary or natural subject matter**: this is the pre-iconographic or factual level in which objects and events are identified (e.g., this is a painting of 13 long-haired men in robes gathered around a long table for a meal).

(2) **The secondary or conventional subject matter**: this is the iconographic level, in which some cultural knowledge of themes and concepts manifested in stories, images, and allegory is needed (e.g., this is not *just* an image of 13 men gathered for a meal, it is a representation of the Last Supper).

(3) **The intrinsic meaning or content**: this is the iconological level, in which the work is interpreted, based on an understanding of the "basic attitudes of a nation, a period, a class, a religious or philosophical persuasion—unconsciously qualified by one personality [the artist] and condensed into one work"[14] (e.g., this painting is Leonardo da Vinci's *Last Supper* from 1498, a mural in the Convent of Santa Maria delle Grazie in Milan, Italy. It depicts the internal confusion of the 12 disciples after Jesus announced one of them would betray him—each one wondering if it would be him). On this level, meaning depends upon an understanding of the two previous levels. This level requires a sophisticated understanding of world cultures, symbolism, and the significance of the work and its context in art history.

Nearly 80 years later, Panofsky's categories are still useful in understanding the ways in which the content of visual images can be analyzed. With art works it is certainly easiest to describe at the pre-iconographic level, that is, to enumerate objects and scenes represented in the image, rather than any literary themes or intrinsic meaning. In some cases, it may be possible to identify a subject concept (e.g., a depiction of a battle scene) but also to determine from a work's title a specific instance of the concept (e.g., the Battle of Gettysburg). LIS scholar Sara Shatford Layne relates Panofsky's first level of meaning to the *of-ness* of the resource (i.e., what this is an image *of*) and his second level to *aboutness*; she states that the third level cannot be used to analyze visual images with any degree of consistency.[15] With musical works, it is even harder to identify concepts or to enumerate what themes and topics are being represented. If one wants *true* conceptual analysis of nontextual information resources, such analysis would be at the interpretive thematic level. It is fairly easy to describe how an object looks, but the identification of intrinsic meaning or iconological significance for a nontextual resource requires special study and training.

Exhaustivity

When examining documents for subject content, one must have a clear idea about the level of exhaustivity that is required. *Exhaustivity* is the number of concepts that will be considered in the analysis. The number of concepts included in the analysis and the subject description may be guided by any of the following:

- local policy (e.g., use no more than three main concepts),
- type of material (e.g., archival collections require more depth of indexing), or
- published guidelines for translation (e.g., the *Subject Headings Manual* limits the number of LCSH terms that may be applied to a resource).

A. G. Brown identifies two basic degrees of exhaustivity: *summarization* and *depth indexing*.[16] Summarization identifies only a dominant overall subject of the resource, recognizing only concepts embodied in the main theme. Depth indexing aims to extract all the main concepts addressed in a resource, recognizing subtopics and lesser themes.

In library cataloging, subject analysis has traditionally been carried out at the summarization level, reserving depth indexing for other enterprises such as archival description or periodical indexes. That is, in the cataloging of books and serials in libraries, the cataloger generally attempts to find the overall subject concepts that encompass the whole resource. Depth indexing has traditionally been reserved for parts of resources (e.g., articles in journals, chapters in books) and has usually been done by commercial indexing enterprises. In the case of an open access journal on the web, such as the *Journal of Statistics Education*,[17] the subject at the summarization level can be no more in depth than *statistics education,* even though the subjects of individual articles are much more specific.

In libraries, whole resources *could* be indexed more deeply, but it raises questions about how accurate, precise, or helpful the resulting descriptions would be. One must consider the users' needs. If a 900-page resource contains only a few paragraphs or a few pages on a particular topic, is it worth the users' time to be directed to that resource? For the example we used earlier, *Introduction to Cataloging and Classification*, a deeper look at the table of contents shows the various other concepts that *might* be included if this resource were indexed at a more granular level of exhaustivity (see Figure 9.3). Instead of just the summarization level subject headings enumerated earlier in the chapter (**Descriptive cataloging**, **Subject cataloging**, **Classification—Books**, and **Resource description & access)**, additional terms representing *MAchine-Readable Cataloging (MARC) format*, *original cataloging*, *authority control*, *authority files*, *system design*, *subject analysis*, *subject headings*, and more would be added. Although portions of the resource are about these concepts, the library has generally focused on pointing users to resources that are entirely or primarily about the topics listed as subject headings in the metadata. The smaller concepts are subsumed under the broader ones (e.g., *original cataloging* and *authority control* are considered to be components of **Descriptive cataloging**; *subject headings* and *subject analysis* are parts of **Subject cataloging**).

Figure 9.3. A Detailed Page from the Table of Contents in *Introduction to Cataloging and Classification*.

It should be pointed out that many books (though certainly not all) have extensive back-of-the-book indexes, and this has been one of the justifications for subject cataloging at the summarization level. That is, there is a difference in degree between *document retrieval* and *information retrieval* (see Figure 9.4).

Summarization allows for document retrieval, after which many users consult the document's internal index (which, for electronic resources, may mean conducting a word search within the text) to retrieve the relevant information they need from the document. Depth indexing, however, allows retrieval at a much more specific level, even to the retrieval of sections or paragraphs in a document.

Exhaustivity affects both precision and recall in retrieval. *Precision* is the measurement of how many of the documents retrieved are relevant. *Recall* is the measurement of how many of the relevant documents in a system are actually retrieved. Depth indexing is likely to increase precision because more specific terminology is used. Summarization is likely to increase recall because the search terms are broader and more sweeping in their application of terminology.

The summarization approach is very useful in retrieving tangible resources (e.g., books, DVDs, print journals). With intangible digital resources (e.g., online zines, datasets, websites), we must think carefully about summarization versus depth indexing. Search engines do the ultimate depth indexing (although through automatic rather than intellectual means). Often, the occurrence of a word anywhere in an online resource means that the site will be retrieved by a search on that word, whether or not the word reflects any topics actually covered in the resource. However, because this retrieval approach is based on matching specific strings of characters, rather than searching for the concepts signified by those strings, depth indexing increases recall while greatly decreasing precision.

Figure 9.4. Continuum of Exhaustivity.

In addition to deciding the level of exhaustivity, we also face the problem of deciding what is an analyzable unit. Traditionally, whole resources have been the analyzable units in libraries, collections are described in archives, individual works of art are depicted in museums, and individual articles are the units described in commercial indexing enterprises (along with individual poems, stories, or essays published in collections). In digital libraries and other online spaces there is, so far, no definition for an analyzable unit. Should it be the whole website or individual pages? Should it be a single image, the entire collection of 500 images scanned for a digital library project, or some combination of both? Once we have decided what is being described, we will have to rethink whether our users need summarization-level subject headings or depth indexing vocabulary for the analyzable units.

Objectivity

The challenges addressed above (levels of complexity, difficulty, consistency, etc.) raise questions as to whether the subject analysis process can be an objective or impartial one. For example, some find it much easier to analyze resources about topics with which they are already familiar or those that they simply enjoy; some find it difficult to analyze resources they dislike or disagree with on a philosophical, moral, social, religious, or political basis. Information professionals are expected to remain objective and impartial in all of their work-related activities, but is this realistic? Daniel N. Joudrey found that 75 percent of the LIS students who participated in his study of aboutness determination questioned the validity of the creator's premise in at least one of the three information resources they were analyzing.[18] Is this surprising? Information professionals (and LIS students) are only human after all. Although information professionals are expected to remain neutral, there is a human tendency to judge the information that we encounter, whether it is a bad first impression from cover art or lasting doubts about some creators' evidence to support their claims. With some controversial resources or resources representing an opposing point of view, judgments and preconceptions may be unavoidable. In such cases, it is important to be aware of and to acknowledge one's biases, prejudices, and beliefs when conducting the conceptual analysis, and to seek the opinions of others when needed.

Those with vastly different understandings of the world might vehemently disagree as to how objective the subject analysis process is or can be. While some with a more positivist, empirical view of the world might believe that there is an innate, identifiable aboutness in each information resource just waiting to be discovered, others with a more constructivist understanding may view the process as one that can only be performed through the lens of the analyst's own background, knowledge, culture, responsibilities, and even mood. In other words, they see conceptual analysis as a highly subjective, interpretive process that is dependent on human skills of observation, interpretation, and analysis. Whatever one's epistemological orientation, the work of subject analysis must provide subject access to information resources in a systematic, deliberate fashion. Consequently, information professionals often forego long philosophical debates over the nature of reality, aboutness, and subject determination, and just do the task—with an understanding that although cataloging is not a neutral act, we should attempt to keep our biases in check as much as possible while performing the process and remember that self-awareness is crucial.

SOME METHODS USED TO DETERMINE ABOUTNESS

As stated earlier, not everyone agrees on how to approach the determination of aboutness, so there is no single process that is used by everyone. Over the years, various methods have been offered. Some were simply lists of bibliographic features that could be consulted, whereas others provided questions and concepts to consider while examining a resource. Some view the process as comprising two activities, while others believe there are three or four discrete steps. Some do not think of the process in terms of steps at all. In this section, some of the most familiar and useful approaches to determining aboutness are discussed.

Langridge's Approach

Langridge views the subject analysis process as a series of discrete activities.[19] He also stresses that the conceptual analysis must be performed independently from any particular classification scheme or controlled vocabulary, and that the analysis should be written down so as to avoid muddled notions of aboutness.[20] In addition to examining the various parts of the text, he states that the information professional must keep three basic questions in mind in order to determine the aboutness of an information resource. Those questions are:

- What is it?
- What is it for?
- What is it about?[21]

According to Langridge, the first question is answered by one of the fundamental forms of knowledge. In other words, one must ask: Is this resource a work of history? Is it science? Is it philosophy? Is it art? Langridge identifies 12 distinct forms of knowledge into which all resources fit:

- Philosophy
- Natural science
- Technology
- Human science
- Social practice
- History
- Moral knowledge

- Religion
- Art
- Criticism
- Personal experience
- Prolegomena (logic, mathematics, grammar—the foundations of knowledge)[22]

Langridge's second question looks at the purpose of the document. Why was it created? How might it be used? Looking at disciplines may help to answer the second question. Is it for a specific audience? Is this book on cows meant for a zoologist? Is it for a veterinarian? Is it for a dairy farmer? Is it for a child, or a chef? Is it for a vegetarian, or a researcher studying the place of cows in spirituality? These reflect several different perspectives that might affect our understanding of the resource and its aboutness. Langridge sees disciplines as ever-evolving areas of interest or specialization falling under the 12 forms of knowledge.

The answer to the third question is one or more topics. Topics are the everyday phenomena that we perceive, that is, concepts, objects, places, events, people, groups, and so on. Langridge points out that topics are not specific to any one form of knowledge or discipline. For example, the topic *clothing* can appear in any number of disciplines or sub-disciplines, such as clothing design, clothing manufacturing, home economics, social customs, psychological or religious aspects of clothing, and so on. Langridge, in his approach to subject analysis, also includes some other characteristics as part of the process, such as examining the nature of the text (bibliographic structures and mediums of communication) and the nature of the thought (point of view, type of writing, audience, intellectual level, etc.); these help to bring out additional aspects or perspectives of the aboutness.

Wilson's Approaches

Wilson has described four possible methods that one may use to come to an understanding of what a resource is about.[23] The first might be called the *Purposive Method.* In this approach, one tries to determine the creator's aim or purpose in creating the information resource. If the creator gives a statement of purpose, then we may presume to know what the work is about. But some creators give no such statement, others seem to aim at several things at once, and others provide multiple statements indicating different purposes or objectives in each.

Wilson's second method of deciding what a resource is about might be called the *Figure-Ground Method.* Using this method, one tries to determine a central figure that stands out from the background of the rest of the information resource. This might be an idea, a person, an object, a place—whatever is the central aspect of the subject. What stands out, however, depends on the observer of the resource as well as on its creator. What catches one's interest is not necessarily the same from person to person, and may not even be constant for the same person a few months or several years later. Education, background knowledge, intellectual interests, and initial assumptions about the information resource can also influence what stands out.

Wilson's third method is the *Objective Method* (which, by the way, is the method used in most attempts at automated conceptual analysis). One tries to be objective by "counting" references to various concepts to determine which ones vastly outnumber the others. Unfortunately, something constantly referred to in the resource might be a background idea (e.g., Germany in a work about World War II). This method is also difficult because a primary concept might be signified by different words throughout the resource. For this method to be successful, the information professional must understand that the variously expressed concepts are related. It is also possible that the concept that is central to the aboutness might not be expressed explicitly in the text. Wilson gives the example of a work being about a person's political career without those exact words ever appearing in the text. Collantes found that when people were asked to read abstracts and then write down subject words or phrases that they believed conveyed the meaning in the abstracts, 8 percent of the readers used words that did not appear anywhere in the abstract.[24]

The last of Wilson's methods is the *Cohesion Method,* an approach that looks at the unity of the content. When using this method, one tries to determine what holds the work together, what content has been included, and what has been left out of the treatment of the topic. Again, the observer of the information resource has to be objective and has to know quite a lot about the subject in order to recognize what was omitted. In addition, there may be several ways in which the work can appear to be unified, and creators do not always reach the ideal of a completely unified presentation.

Use-Based Approaches

Over the years, a number of other approaches to subject analysis have been discussed in the LIS literature. Some can be described as *use-based* approaches. The main idea is that aboutness can be determined by looking at how a resource could be used or what questions a resource could answer. Lancaster, concerned about users and how the resource might be used, suggests asking three questions to determine aboutness:

- What is it about?
- Why has it been added to our collection?
- What aspects will our users be interested in?[25]

Concerns about use, purpose, or patrons' interests, while informative, will not suffice as the only approach to aboutness determination because the information professional is being asked to perform an impossible task: to predict all of the possible (present and future) uses for a document. No matter how skilled one may be, this approach amounts to little more than guessing why an information resource may be needed. Therein lie the difficulties associated with use-based approaches to aboutness. This is not to say that asking how a document could be used has no place in aboutness determination. On the contrary, it is a very important question; it just cannot be the *only* question.

As this discussion indicates, there seems to be no one correct way to determine aboutness. Different approaches to identifying relevant aboutness data may be used by different analysts, or by the same analyst with different resources or on different days. One can use any or all of these methods, but the different methods will not necessarily lead to the same result. If they give the same result, as they often could, it would appear that the subjects have been identified. However, a single person might arrive at three or four different subjects using different methods, and several persons might arrive at different results using the same method. In the following section, the authors provide an approach to conceptual analysis that takes into account some of the important components described above, along with some insights from Joudrey's research on the topic.

CONCEPTUAL ANALYSIS PROCESS

Conceptual analysis comprises three interconnected components:

- an examination of the physical resource (or the display of a digital resource),
- an examination of the intellectual or creative content, and
- numerous simultaneously performed stages of aboutness determination.

In the following sections, these components are addressed.[26]

Resource Examination

The conceptual analysis process begins with an examination of the information resource, particularly its bibliographic features and visual elements. Similar to the process of descriptive cataloging, in which both content and carrier are considered when creating metadata, the conceptual analysis of a resource requires an examination of both the intellectual content and the physical resource (if the resource is in a tangible form). The examination starts with the parts of the resource that stand out. In many instances this begins with the information resources themselves, but in other instances it begins with accompanying materials (e.g., manuals, containers, inserts, cases, labels). Somewhat different techniques have to be used for information resources that contain text versus those that contain nontextual information. Concentrating first on the textual resources, the following parts should be considered:

- **Cover, Jacket, or Container**: Generally, the cover (or jacket or container) is the first thing one sees when examining a tangible resource. It kicks off the input process. Without even realizing it, a person seeing the cover gets a first impression of the resource. For that reason alone, this source must be considered. This may be where the title information and the creator's name are first encountered (which may or may not be identical to the information found on the title page or other preferred sources of information). In addition to basic information about title and creator, one may encounter a great deal of visual information from this source. Whether this information is helpful depends upon many factors. A cover may be explicitly designed to include meaningful imagery or it may be designed simply to catch the eye; its illustrations may be subtle, symbolic, literal, apparently random, or simply a clever marketing strategy. While there are no guarantees that this information will be useful, it should be considered a potential source of aboutness data.

- **Title and Subtitle**: A title can be helpful in giving an immediate impression of the topic of a document, but titles can also be misleading. Occasionally, more than one form of title or subtitle may be found on an information resource. The title *Introduction to Cataloging and Classification* is quite straightforward. On the other hand, the title *A Compendium of Tiddlywinks Perversions* is not so clear and is not assisted by its other title information, *Alleghany Airlines Book Club Presents.*[27] Another example, *What the Thunder Said,* turns out to be a website devoted to the life and works of T. S. Eliot.[28]
- **Table of Contents**: A list of contents can help to clarify the main topic and identify subtopics. It can be especially helpful for resources that are collections of articles, papers, reports, and the like, by different creators. Tables of contents can show the variety of specific topics covered in a resource. Individual chapter titles, like titles proper, are not always helpful or clear, though. Some creators prefer to use ironic, tongue-in-cheek, or cryptic chapter titles; so, the table of contents is only one of the sources consulted during the conceptual analysis.
- **Introduction or Preface**: Introductions, prefaces, opening essays, or some equivalent feature may contain some of the most useful, concrete, and straightforward aboutness data in the entire resource. Unfortunately, these features are not always present in resources. Often, an introduction is an aid in determining the creator's overall purpose or objective in creating the work (as suggested by Wilson[29]) and may also serve to indicate a creator's point of view or perspective on the topic (as recommended by Langridge[30]). For example, the introduction to *JGarden: The Japanese Garden Database* (see Appendix A) explains that this resource is not just descriptions of the gardens but draws upon literature, paintings, images, and so forth, to provide "comprehensive visual and textual information on the history, construction, materiality, people, language, patterns and processes by which these gardens were constructed."[31]
- **Illustrations, Diagrams, Tables, and Their Captions**: Illustrations and their captions are particularly important in assessing subjects in fields such as art, where in many cases, illustrations make up the vast majority of the content and therefore must be examined in order to determine aboutness. The captions for illustrations often are quite descriptive of subject content. In some resources, however, the illustrations may provide little helpful aboutness information or may even be distracting.
- **Other Bibliographic Features**: Some information professionals also consult dedications and acknowledgements, hyperlinks, abstracts (if present), and indexes. These elements may confirm or contradict impressions gained from examining the title, table of contents, introduction, and so on. Similar to Wilson's objective method, a back-of-the-book index can show what topics are given the most attention by showing the number of pages devoted to each;[32] what it does not provide, however, is context to show which topics are related and how they are related.[33]
- **The Text**: In addition to examining the bibliographic and visual features, the text itself should be examined to get a more complete understanding of the information resource. In looking through the text, one may encounter helpful information in the introductory sections, opening paragraphs, conclusions, chapter summaries, and bolded section headings throughout a resource. Section headings, if used, can provide a quick overview of the content found in individual parts of a chapter. There are several approaches that may be taken to examining the text, the most common being *skimming.* Skimming allows the analyst to get a sense quickly of what the creator has written and of the topics covered in the document. This process can be used to help reinforce, refine, or refute emerging aboutness assumptions. In addition to skimming, some may choose occasionally to sample paragraphs or to read longer passages when needed. While examining the text, some may choose to work in a linear fashion from the beginning to the end of the resource, while others may choose to focus only on the beginning *and* the end of the information resource, under the assumption that the richest aboutness data will be found in the introductory and concluding

chapters, sections, or features. Still others may flip randomly through a resource stopping only at what catches the eye. Individuals must find the approach that works best for them and best for the resource they are examining at that moment. No single approach will work for every resource equally well every time.

- **Nontextual Information**: For tangible nontextual resources, one has to examine the object, picture, or other representation itself. Some nontextual resources are manufactured commercially and include accompanying materials such as boxes with text, booklets, instruction sheets, labels, and so on. Intangible electronic resources that are basically images or other forms of artistic work quite often have captions that explain something about them. For individual works or objects with no accompanying text, however, one must examine the resources themselves and translate those ideas into words (which can be quite difficult, if not impossible, without special training or education).

Content Examination

During the examination of the physical resource, one must also be concerned with various aspects of the intellectual and creative contents of the resource. In this section, identifying concepts of interest, important content characteristics useful in aboutness determination, and some content examination strategies are addressed.

Identification of Concepts

Different types of concepts can be used as subjects of information resources, including: topics, several types of names (e.g., persons, corporate bodies, geographic areas, titles of works), and time periods.

Topics Used as Subject Concepts. Most people think of topical terms when asked to identify the subject of a resource. A topic can represent a principle object of attention or a theme running through an information resource. Topics can be concrete or abstract. Ideas, objects, phenomena, activities, processes, structures, groups, substances, and more, can be the topics of works. In short, an information resource can be about anything imaginable.

Names Used as Subject Concepts. In the process of determining what a document is about, it may be found that the subject, or one aspect of the subject, is a person, a corporate body, a geographic area, a title, or some other named entity.

- **Persons**: An individual person, living or dead, can be the topic of a book, a movie, or some other form of information resource. It may be primarily biographical or it may cover aspects of a person's career (e.g., the aforementioned site dedicated to T. S. Eliot). Such a work is also, in a sense, about one representative of a group of persons (e.g., literary authors). The Library of Congress, in applying LCSH, for example, has the policy of assigning a subject heading for a group to which a person belongs, as well as for the individual person, on the assumption that if an information seeker wants to learn about *literary authors,* then a site about one such person may be of use. A work can be about deities, biblical figures, legendary and fictitious characters, and named animals as well.
- **Corporate Bodies**: A corporate body is an organization or group of persons who are identified by a name and who act as an entity. A corporate body may be the subject of a resource about an organization, a musical group, a company, a legislative body, a church group, a vessel, and so forth (e.g., a book may be about Exxon-Mobil, The Beatles, or the Cambridge Public Library). There are also entities whose names resemble corporate bodies, but they are not. Sometimes such bodies have the same name as the building they work in. This is often true of churches, for example, and then one has to be certain whether the building or the corporate group is the topic of the work.
- **Geographic Names**: Geographic names can take different roles in the determination of subject content. In some cases, a document may actually be about a specific place (e.g., a work about the history of Brookline, Massachusetts). However, much of the time, a work is not about the place *per se,* but the geographic location provides a context for the topical content, as in a work about the projects and life of architect Julia Morgan, who did virtually all of her work in California. Falling between these extremes is the case where the geographic area is the *subject* of the topic. An example is the exhibition catalog, *"A Sweet Foretaste of Heaven": Artists in the White Mountains, 1830–*

1930.[34] The exhibition consisted of landscape paintings of the White Mountains of New Hampshire.

- **Titles**: Titles are another type of entity that may be used as a subject concept. A resource can be about one or more scholarly, literary, musical, or artistic works. Reviews, adaptations, critical essays, histories, parodies, and the like are some examples of the types of resources that might reference the title of another work in its subjects.
- **Other Named Entities**: Some named entities resemble both corporate names and geographic names but are neither. Entities such as named buildings, structures, cemeteries, bridges, and archaeological sites fall into this category. An example is the subject heading **Megiddo (Extinct city)**, which is used to represent the archaeological site.

Chronological Elements. The time period can be an important aspect of the subject content of information resources. Time periods limit the coverage of the topic and therefore dictate content in subtle ways. For example, resources about computer access to information in the 1970s will not include information about the Semantic Web or Open Access journals. Time can be expressed in a number of ways. Named periods (e.g., Bronze age) and styles (e.g., Rococo) often act as surrogates for chronology. These are of particular importance in the fields of art, architecture, music, history, and literature, but they have not been particularly well handled by controlled vocabularies or by the MARC format. Only specific dates or date ranges are usually treated as separate chronological elements, because named periods and styles generally have been treated as adjectival qualifiers to topical information.

Content Characteristics

Just as topics, names, and chronological elements are important to understanding the aboutness of information resources, so are some additional characteristics that are related to the content and some that are related to the creator. These characteristics include what Langridge discussed as "the nature of the text" and "the nature of the thought."[35] According to Langridge:

> There remain a number of very important characteristics requiring identification which have always been treated as part of the process of subject analysis. I shall refer to these as formal characteristics to distinguish them from the real subject features. Though none of these formal elements alters the subject of a document, some of them can make a considerable difference to its treatment or presentation.[36]

For some information resources, a few of these content characteristics may be relevant in the conceptual analysis process, but with other resources, perhaps none will be useful at all. Like so many other factors in the conceptual analysis process, it will depend on the nature of the resource being analyzed. The following characteristics should be considered if they are applicable.

Research Methods. Although this characteristic will not be appropriate for most popular or creative works, some academic research materials, scientific articles, technical reports, and the like might benefit from considering the methods used by the researcher to examine the issue, the hypothesis, or the research questions being addressed in the resource. This characteristic might be useful in helping one to better understand the aboutness, the form, and the level of the work, as well as the means used to reach the conclusions. It also may be particularly useful in the indexing of scientific datasets, papers, and articles.

Point of View. Except for Langridge, point of view is rarely mentioned in discussions of the conceptual analysis process. Although not all resources have a specialized or identifiable point of view, some of them are written from a particular perspective, which could be of special interest to some users and could be anathema to others. This content characteristic can be helpful when analyzing some types of resources and necessary when describing those works from a particular viewpoint. For resources that may be controversial in some fashion (e.g., political works; religious works; cultural treatises; works on sexuality, gender, age, or socio-economic levels), the point of view may be an important piece of metadata that will help the users to find, identify, and select the resources that they need. This content characteristic, however, is rarely if ever translated into controlled vocabulary terms or classification notations; instead, it may be addressed in summary statements or abstracts.

Language, Tone, Audience, and Intellectual Level. These four characteristics are addressed together because they are in many ways interrelated. The relationships among them are multifaceted and myriad. For example, the language chosen by the creator helps to set the tone of the work. The language and tone used are shaped by the intellectual level that the creator is striving for and by the creator's intended audience for the work. In examining the work, the language and tone are particularly helpful in understanding the audience and intellectual level (e.g., whether it is an academic or a popular work). While none of these directly affects aboutness, they may help to establish some context for understanding the information resource as a whole.

These characteristics may also affect the resource examination process. The analyst may need to treat a scholarly intellectual work differently than a popular work (e.g., a scholarly resource may need more exploration of the content, more reading of the introduction and conclusion, and more time to complete the analysis than does a work of popular culture). Of those writing on aboutness determination, few address these issues. Langridge addresses one of the four—*audience*—in his discussion of "form of thought." Audience may also appear as the answer to his conceptual analysis question: *What is it for?*[37] Joudrey found that 92 percent of the participants in his qualitative study of aboutness were concerned with audience. Although the study participants found *audience* helpful in the aboutness determination process, some were reluctant to include the concept in their final descriptions of the resource. He notes, "the participants included audience [in their aboutness statements] when the book was written for and directed toward one particular audience, and they excluded the concept when the item could appeal to multiple audiences or a more general audience."[38]

Language and tone are almost never addressed in the subject analysis process. Audience and intellectual level are sometimes addressed in the translation stage of subject analysis (e.g., in subject headings), but audience may also appear in other places in metadata descriptions (e.g., in MARC bibliographic records, headings from *Library of Congress Demographic Group Terms* are placed in the 385 field for audience characteristics; *Audience* is used as a supplemental element in some implementations of Dublin Core).

Genre and Form. The final characteristic to consider in the conceptual analysis process is the *genre* or *form* of the information resource being analyzed, or the form of important parts of that resource. Genre and form are not strictly subject features. They describe what a resource *is* rather than what it is about. They are concepts, however, that have been associated with subject analysis from the inception of the idea that books could be entered in catalogs and placed on shelves according to categories to which they belonged. Early categories included such forms as encyclopedias, biographies, and histories, as well as subjects such as chemistry and religion. Later, as subject headings evolved to mean what a resource is about instead of a category to which the book belonged, the idea of form and genre remained as part of the subject analysis process. Because it was often difficult to separate the idea of genre/form from aboutness, as in the case of *history* (which seems to incorporate elements of both), the concept of genre/form has only recently begun to be treated differently in metadata descriptions.

In an effort to aid in the process of separating genre and form from subject, the Subject Analysis Committee (SAC) of the American Library Association (ALA) devised a definition for form. The appropriate ALA bodies officially approved the definition in January 1993:

> Form data are those terms and phrases that designate specific kinds or genres of materials. Materials designated with these terms or phrases may be determined by an examination of:
>
> - their physical character (e.g., videocassettes, photographs, maps, broadsides)
> - the particular type of data that they contain (e.g., bibliographies, questionnaires, statistics)
> - the arrangement of information within them (e.g., diaries, outlines, indexes)
> - the style, technique, purpose, or intended audience (e.g., drama, romances, cartoons, commercials, popular works) or a combination of the above (e.g., scores)
>
> A single term may be modified by other terms, in which case the whole phrase is considered to be form data (e.g., aerial photographs, French dictionaries, conversation and phrase books, wind ensemble suites, telephone directories, vellum bound books, science fiction).[39]

In the past decade, LC has been paying special attention to this content characteristic, going so far as to create a new thesaurus of terms just for genre/form concepts. Its introduction states,

> Genres and forms may be broadly defined as categories of resources that share known conventions. More specifically, genre/form terms may describe the purpose, structure, content, and/or themes of resources. Genre/form terms describing content and themes most frequently refer to creative works and denote common rhetorical devices that usually combine elements such as plot and setting,

character types, etc. Such terms may be closely related to the subjects of the creative works, but are distinct from them.[40]

Separating genre/form from subject became increasingly impor-tant as retrieval systems became more advanced and the organizing world became more attuned to organizing information that is not just in textual form. In the discipline of music, for example, identifying genre/form has always been critical and has been accommodated by treating it as subject. Now more users are looking for other kinds of information (e.g., city ordinances, chalk drawings, digital maps, sculpture reproductions). By separating genre/form from subject, it is possible to take advantage of system design to allow searching for resources in specific genres or forms. Some metadata schemas have elements that are defined specifically for this concept (e.g., *Form of work* in RDA; *Type* and *Format* in Dublin Core; *worktype* and *material* in VRA Core).

Content Examination Strategies

As discussed earlier, different approaches to identifying relevant aboutness data may be used by different analysts, or by the same analyst with different resources or on different days. Some may use Langridge's or Lancaster's questions, while others may use one or more of Wilson's approaches, and still others may use no organized method at all. Answering a series of questions, either formally or informally, can help one go through the subject analysis process quickly. A sample set of questions and an example of the process is found in Appendix A of this textbook.

Joudrey found in his aboutness determination research that LIS students with no previous experience or coursework in organizing information mostly used the same approaches to conceptual analysis. The most commonly used approaches are three of the four described by Wilson: *the Purposive Method, the Figure-Ground Method,* and *the Objective Method.* The participants rejected some of the other strategies described in the LIS literature by either using them infrequently (or not at all) or by stating that they were not approaches they had considered or they would consider. For example, although the participants explored the creator's purpose in writing a document, on the whole they did not consider how patrons would use the documents, what questions the documents would answer, or how someone would search for the documents. Based on Joudrey's research, it appears that Patrick Wilson got it right half a century ago.[41] Joudrey concludes:

> Because of the interpretive nature of all of the processes that the participants used (and those that they did not use), it is recommended that each content examination strategy should be used in conjunction with other strategies. Wilson was accurate when he stated that each of the approaches that he described was *a* method, and not *the* method, to analyze aboutness. No one approach to the content was observed in isolation, nor should they be used in isolation. Similar to conducting good qualitative research, the content examination strategies need be triangulated. Getting as many perspectives as possible on the items is the best strategy for determining aboutness.[42]

Stages in Aboutness Determination

In an attempt to better understand what happens during the conceptual analysis process, Joudrey analyzed the activities that were being performed while his study participants were attempting to determine what the information resources were about. While the physical resources were being examined and the participants were considering things like the creator's purpose and what topics were mentioned most often, they were also advancing through a number of simultaneous stages. The researcher has categorized the stages as the *input process*, *assumption making*, the *revision process*, *sense making*, and *stopping*.

The conceptual analysis process begins with an input phase in which data is collected by encountering content in some form or manner (seeing, noticing, envisioning, etc.). This may occur through simple visual examination, more in-depth exploration of general content, or through seeking specific desired chunks of data. Shortly after the input process begins (in some cases as early as viewing the cover), a second process begins in which assumptions about the resource's aboutness are made. These assumptions may be about macro-level, micro-level, or chapter-level aboutness, or they may be about other characteristics of the resource. These assumptions then undergo a revision process, in which assumptions are refined, reinforced, and/or refuted. Concurrently with the input, assumption making, and revision processes, the multifaceted process of sense making begins.[43] This entails a number of individual activities, including finding context, interpreting, comparing, and reasoning. After the first moments of the input process, all of these other processes are performed simultaneously and continuously until an understanding of the resource's aboutness is reached. The final process centers on how and when one decides to stop the examination of the resource.[44]

NEXT STEPS IN SUBJECT ANALYSIS

Once the conceptual analysis is complete, it can be very helpful to write an aboutness statement that begins, "This resource is about . . ." It may be one or two sentences or it may be a short paragraph. As Langridge states, the conceptual analysis should be written down to avoid muddled notions of aboutness.[45] It helps the information professional to keep track of all of the major concepts and to relate them to each other. Depending on the degree of exhaustivity that is being followed, the concepts included may be limited to the main idea of the entire work or to all of the major subtopics found in the resource.

Once the aboutness statement is complete, the aboutness concepts that have been identified must then be translated into the subject languages being used. The information professional identifies the key concepts and chooses terms from the aboutness statement to be searched in the controlled vocabulary. Specific rules for using controlled vocabularies are found in their introductions, as well as, in some cases, manuals that accompany them (e.g., LCSH's *Subject Headings Manual*). In order to translate an aboutness statement into classification notations, it is necessary to understand the hierarchy or the facets of the classification scheme that is to be used. If a hierarchical scheme is used, it is helpful to determine the discipline into which the information resource falls. For example, a history of Spain is usually considered to fall within the discipline of history and the sub-discipline of European history, while a history of chemistry is generally thought to fall within the broad area of the sciences, specifically in the discipline of chemistry. Creating a hierarchical string for discipline, sub-discipline, topic within discipline, accompanied by concepts of treatment, place, time, and form is helpful in translating the aboutness into a hierarchical classification. Using a faceted classification is much like choosing subject terms from a controlled vocabulary that allows terms to be combined into a complex subject heading string. Notations for separate facets must be found in the classification and then must be put together according to the rules for the scheme.

CONCLUSION

This chapter has discussed the nature of *subject* and the process of determining the aboutness of information resources. Determining aboutness must precede the use of the specific subject access tools employed for assigning controlled vocabulary and classification notations. Before being able to make the best use of the specific tools, it is helpful to understand controlled vocabularies and classification schemes in general. The following two chapters address these issues.

Some Important Terms in This Chapter		
Aboutness	Exhaustivity	Precision
Aboutness Statement	Form	Recall
Conceptual Analysis	Genre	Subject Analysis
Controlled Vocabulary	Is-ness	Subject Language
Depth Indexing	Of-ness	Summarization

Some Important Acronyms in This Chapter	
AAT:	*Art & Architecture Thesaurus*
ALA:	American Library Association
DDC:	*Dewey Decimal Classification*
LC:	Library of Congress
LCC:	*Library of Congress Classification*
LCGFT:	*Library of Congress Genre/Form Terms*
LCSH:	*Library of Congress Subject Headings*
LIS:	Library and Information Science
MARC:	MAchine-Readable Cataloging
SAC:	Subject Analysis Committee
UDC:	*Universal Decimal Classification*

NOTES

All URLs accessed June 2017.

1. Working Group on the Future of Bibliographic Control, "On the Record: Report of the Library of Congress Working Group on the Future of Bibliographic Control," Library of Congress, http://www.loc.gov/bibliographic-future/news/lcwg-ontherecord-jan08-final.pdf.

2. F. W. Lancaster, *Indexing and Abstracting in Theory and Practice,* 3rd ed. (Champaign: University of Illinois, Graduate School of Library and Information Science, 2003), 9.

3. Sometimes referred to as *is-ness*, when used in opposition to *aboutness*.

4. Patrick Wilson, "Subjects and the Sense of Position," in *Two Kinds of Power: An Essay on Bibliographical Control* (Berkeley: University of California Press, 1968), 69–92. Also reprinted in *Theory of Subject Analysis: A Source-book,* eds. Lois Mai Chan, Phyllis A. Richmond, and Elaine Svenonius (Littleton, CO: Libraries Unlimited, 1985), 309–20.

5. Richard E. Nisbitt, *The Geography of Thought: How Asians and Westerners Think Differently . . . and Why* (New York: Free Press, 2004).

6. George Lakoff, *Women, Fire, and Dangerous Things: What Categories Reveal about the Mind* (Chicago: University of Chicago Press, 1987), 24–26.

7. D. W. Langridge, *Subject Analysis: Principles and Procedures* (London: Bowker-Saur, 1989), 4.

8. Oliver L. Lilley, "Evaluation of the Subject Catalog: Criticisms and a Proposal," *American Documentation* 5, no. 2 (1954): 41–60.

9. Lourdes Y. Collantes, "Agreement in Naming Objects and Concepts for Information Retrieval" (PhD diss., Rutgers University, 1992), 154.

10. For example: Regene C. Ross, Chair, Task Force on Copy Cataloging, *Report of the Task Force on Copy Cataloging,* May 12, 1993, cited by Thomas Mann, in "'Cataloging Must Change!' and Indexer Consistency Studies: Misreading the Evidence at Our Peril," *Cataloging & Classification Quarterly* 23, nos. 3/4 (1997): 37–38.

11. Langridge, *Subject Analysis,* 9.

12. Richard W. Unger, *The Art of Medieval Technology: Images of Noah the Shipbuilder* (New Brunswick, NJ: Rutgers University Press, 1991).

13. Erwin Panofsky, *Studies in Iconology: Humanistic Themes in the Art of the Renaissance* (New York: Harper & Row, 1972), 5–17.

14. Ibid., 7.

15. Sara Shatford [Layne], "Analyzing the Subject of a Picture: A Theoretical Approach," *Cataloging & Classification Quarterly* 6, no. 3 (1986): 45.

16. A. G. Brown, in collaboration with D. W. Langridge and J. Mills, *An Introduction to Subject Indexing*, 2nd ed. (London: Bingley, 1982), frames 48, 51.

17. *Journal of Statistics Education*, http://www.amstat.org/publications/jse/.

18. Daniel N. Joudrey, "Building Puzzles and Growing Pearls: A Qualitative Exploration of Determining Aboutness" (PhD diss., University of Pittsburgh, 2005), 165–71.

19. Langridge, *Subject Analysis,* 73–98, 136.

20. Ibid., 57.

21. Ibid., 8–10.

22. Ibid., 33–37.

23. Wilson, "Subjects and the Sense of Position," 78–88. The authors of this text have added more consistent and descriptive names for each of the methods.

24. Collantes, "Agreement in Naming Objects," 154.

25. Lancaster, *Indexing and Abstracting,* 9.

26. Some of this section appeared in an earlier form in Arlene G. Taylor, "Books and Other Bibliographic Materials," in *Guide to Indexing and Cataloging with the Art & Architecture Thesaurus,* eds. Toni Petersen and Patricia J. Barnett (New York: Oxford University Press, 1994), 101–19, and in Daniel Joudrey's "Building Puzzles and Growing Pearls."

27. Sunshine, "A Compendium of Tiddlywinks Perversions," North American Tiddlywinks Association, http://tiddlywinks.org/rules-of-tiddlywinks/alleghany-airlines-book-club-presents/.

28. Raymond Camden, "What the Thunder Said," http://www.whatthethundersaid.org/.

29. Wilson, "Subjects and the Sense of Position," 78–88.

30. Langridge, *Subject Analysis,* 73–98, 136.

31. "JGarden: The Japanese Garden Database," http://www.jgarden.org/.

32. Wilson, "Subjects and the Sense of Position," 78–88.

33. Joudrey, "Building Puzzles," 127.

34. *"A Sweet Foretaste of Heaven": Artists in the White Mountains, 1830–1930,* with essays by Robert L. McGrath and Barbara J. MacAdam (Hanover, NH: Hood Museum of Art, 1988).

35. Langridge, *Subject Analysis,* 55–57.

36. Ibid., 45.

37. Ibid.

38. Joudrey, “Building Puzzles,” 198.

39. American Library Association, Association for Library Collections and Technical Services, Subject Analysis Committee, “Definition of Form Data,” http://www.pitt.edu/~agtaylor/ala/form-def.htm.

40. Library of Congress, “Introduction to Library of Congress Genre/Form Terms for Library and Archival Materials,” *Library of Congress Genre/Form Terms for Library and Archival Materials*, 2016 edition (Washington, DC: Library of Congress, 2016), GF-iii, https://www.loc.gov/aba/publications/FreeLCGFT/freelcgft.html.

41. Joudrey, “Building Puzzles,” 357–60.

42. Ibid., 361.

43. This is *not* specifically referring to Brenda Dervin’s Sense-Making research methodology, about which more information can be found in Brenda Dervin, “Sense-Making Theory and Practice: An Overview of User Interests in Knowledge Seeking and Use,” *Journal of Knowledge Management* 2, no. 2 (1998): 36–46.

44. Joudrey, “Building Puzzles,” 351–52.

45. Langridge, *Subject Analysis,* 57.

SUGGESTED READINGS

Chan, Lois Mai, Phyllis A. Richmond, and Elaine Svenonius, eds. *Theory of Subject Analysis: A Sourcebook*. Littleton, CO: Libraries Unlimited, 1985.

Chan, Lois Mai, and Athena Salaba. *Cataloging and Classification: An Introduction*. 4th ed. Lanham, MD: Rowman & Littlefield, 2016. Chapter 10: “Principles of Controlled Vocabularies and Subject Analysis.”

Cleveland, Donald B., and Ana D. Cleveland. *Introduction to Indexing and Abstracting.* 4th ed. Santa Barbara, CA: Libraries Unlimited, 2013. Chapter 10: “Subject Analysis and Aboutness.”

Joudrey, Daniel N. “Building Puzzles and Growing Pearls: A Qualitative Exploration of Determining Aboutness.” PhD diss., University of Pittsburgh, 2005. http://d-scholarship.pitt.edu/id/eprint/10357.

Joudrey, Daniel N., Arlene G. Taylor, and David Miller. *Introduction to Cataloging and Classification*. 11th ed. Santa Barbara, CA: Libraries Unlimited, 2015. Chapter 11: “Subject Access.”

Lakoff, George. *Women, Fire, and Dangerous Things: What Categories Reveal about the Mind.* Chicago: University of Chicago Press, 1987. Part I: “Categories and Cognitive Models.”

Lancaster, F. W. *Indexing and Abstracting in Theory and Practice.* 3rd ed. Champaign: University of Illinois, Graduate School of Library and Information Science, 2003.

Lancaster, F. W. *Vocabulary Control for Information Retrieval.* 2nd ed. Arlington, VA: Information Resources Press, 1986. Chapter 1: “Why Vocabulary Control?”

Langridge, D. W. *Subject Analysis: Principles and Procedures.* London: Bowker-Saur, 1989. Chapter 5: “Summarization,” Chapter 6: “Summarizing in Practice,” and Appendix 2: “The Forms of Knowledge.”

Pettee, Julia. “The Subject Approach to Books and the Development of the Dictionary Catalog.” In *Theory of Subject Analysis: A Sourcebook,* edited by Lois Mai Chan, Phyllis A. Richmond, and Elaine Svenonius, 94–98. Littleton, CO: Libraries Unlimited, 1985.

Wellish, Hans H. “Aboutness and Selection of Topics.” *Key Words* 4, no. 2 (March/April 1996): 7–9.

Wilson, Patrick. *Two Kinds of Power: An Essay on Bibliographical Control.* Berkeley: University of California Press, 1968. Chapter 5: "Subjects and the Sense of Position." Reprinted in *Theory of Subject Analysis: A Sourcebook,* edited by Lois Mai Chan, Phyllis A. Richmond, and Elaine Svenonius, 309–20. Littleton, CO: Libraries Unlimited, 1985.

CHAPTER 10

SYSTEMS FOR VOCABULARY CONTROL

In the digital age, subject searching remains a vital approach to finding information resources. Users search catalogs, indexes, digital libraries, and other retrieval tools to find resources that match their interests. In this increasingly web-centric information environment, a search engine is often the first retrieval tool that users approach to find information about a topic they want to explore. Yet some users are frustrated with the millions of results from keyword searches performed using Google and other search engines. With the massive increase in availability of recorded information, it has become more and more evident that keyword searching alone—the approach primarily used by search engines—will not suffice for all information retrieval tools in all information environments. One of the reasons that problems occur is that when keywords are the focus of a search, myriad semantic difficulties can arise. Virtually every common word in the English language has more than one meaning or sense,[1] and many of those senses have more than one nuance; many words can be used as nouns, verbs, adjectives, and/or adverbs. Search systems that purport to allow the use of natural language cannot yet successfully distinguish among different meanings or various parts of speech in very large general systems, although some progress has been made in narrow subject areas.

In addition, there is evidence that people writing about the same concepts often do not use the same words to express them, and people searching for the same concept do not think of the same words to search for it. Many of the inter-indexer consistency studies in the second half of the twentieth century asked participants to think up words in their heads, not to take vocabulary from a list; therefore, although these studies have been used by some authors to "prove" that subject indexing is worth little because indexers are inconsistent, what the studies really showed was that people do not think of the same terms to express the same *aboutness* concepts. Thomas Mann has given an excellent analysis of some inter-indexer consistency studies that support these observations.[2] The clear implication is that controlled vocabulary—usually in the form of a list of authorized terms with their accompanying cross-references—is needed to reconcile all the various possible words that can be used to express a concept and to differentiate among all the possible meanings that can be attached to certain words.

In recent years, however, it has also become clearer that controlled vocabularies are not necessarily appropriate for *all* information environments. Controlled vocabularies appear to work best in contained systems. It is only feasible to create, maintain, and apply them in systems such as indexes, databases, or catalogs, which rely on human expertise for success. A controlled approach to terminology does not and cannot work well for a large-scale, distributed system like the *current* Internet, where no one is responsible for identifying the subjects of works and assigning authorized descriptors, and where the number of resources is enormous.

Several questions are addressed in this chapter. They include: How is subject content expressed verbally in metadata? How are controlled vocabularies structured? Is a controlled vocabulary necessary or can natural language approaches suffice in some environments? How do ontologies assist in identifying concepts? How are tagging and keyword searching used with subject content?

WHAT ARE CONTROLLED VOCABULARIES?

A *controlled vocabulary* is a list or database of terms (or phrases), in which all terms representing a concept are brought together. In other words, it is a *data value standard*—a list of controlled values that can be used to populate a metadata element. In this chapter, we focus mostly on controlled vocabularies that reflect the subjects of information resources, but, in fact, a controlled vocabulary list can contain many types of terms.

For example, a controlled vocabulary might be a list of resource types, languages, media types, demographic terms, or genre/form headings. The term *controlled vocabulary* is another way of expressing the idea of authority control, but usually for entities other than names and titles. As with name authority control, there are three major concerns in establishing a controlled vocabulary: consistency, relationships, and uniqueness.

In a controlled list, usually one of the terms representing the same concept is designated as the *preferred* or *authorized term* to be used in metadata descriptions. Choosing a term as the authorized one is an attempt to control synonyms and nearly synonymous terms. It ensures consistency, which allows for greater collocation of concepts in a retrieval tool. For example, in *Library of Congress Subject Headings* (LCSH), the term **Private investigators** has been chosen for use instead of alternative terms such as *Gumshoes*, *Private detectives*, *Private I's*, or *Private eyes*. Using the preferred term consistently means that all resources about this profession can be retrieved together by searching for that term. The alternatives act as cross-references that point to the chosen term.

Gumshoes
See **Private investigators**

Identifying relationships among terms, therefore, is another key function of some controlled vocabularies (though not all). The network of relationships is referred to as a vocabulary's *syndetic structure* (i.e., its web of interconnections). Relationships among the authorized terms may be identified as *references*, *broader terms*, *narrower terms*, or *related terms*. As seen above, the terms that are not designated as preferred have references from them to the chosen terms or phrases. The authorized term, **Private investigators**, also has other relationships that are specified in the vocabulary. These include

- one broader term—**Police, Private**;
- several narrower terms—**Gay private investigators**, **House detectives**, **Store detectives**, and **Women private investigators**; and
- one related term—**Detectives**.

Even though a word chosen as the authorized heading may have multiple meanings, in a controlled list each term should represent a single concept only. In other words, a controlled vocabulary usually requires homograph control or disambiguation among different meanings of the same word. For example, one controlled vocabulary includes two specific uses of the word *plates*, but each is given a parenthetical qualifier to separate them:

Plates (Engineering)
Plates (Tableware)

Each term in a controlled list should be unique and unambiguous. Definitions, scope notes, creation dates, identifier codes, associated classification numbers, categories, and other features can also assist with this.

TYPES OF CONTROLLED VOCABULARIES

According to *ANSI/NISO Z39.19-2005 (R2010): Guidelines for the Construction, Format, and Management of Monolingual Controlled Vocabularies*[3] (henceforth referred to as Z39.19), controlled vocabularies roughly fall into four major categories:

- **Simple Term Lists**
- **Synonym Rings**
- **Taxonomies**
- **Thesauri** (including Subject Heading Lists)

Each is described in the following sections with examples provided.

Simple Term Lists

A *simple term list* (sometimes called a *pick list*) is not a particularly sophisticated form of controlled vocabulary. It is a straightforward listing of limited values that may be used for a particular metadata element. There is no real concern about semantic relationships in a simple term list. These lists are usually presented in alphabetical or in some other logically evident order (e.g., geographic contiguity, chronological order). Simple term lists tend to be used with *non*-subject metadata, such as format, location, language, and so on, and they lend themselves to being placed in pull-down menus in metadata templates because usually they are finite in scope. An example is the RDA controlled vocabulary for the *media type* element.[4] It provides a straightforward alphabetical list of choices to describe the form of media a resource represents:

- **Audio**
- **Computer**
- **Microform**
- **Microscopic**
- **Projected**
- **Stereographic**
- **Unmediated**
- **Video**

This example also serves as a reminder that not all controlled vocabularies are subject-focused. They are also used in descriptive cataloging, encoding, and so forth.

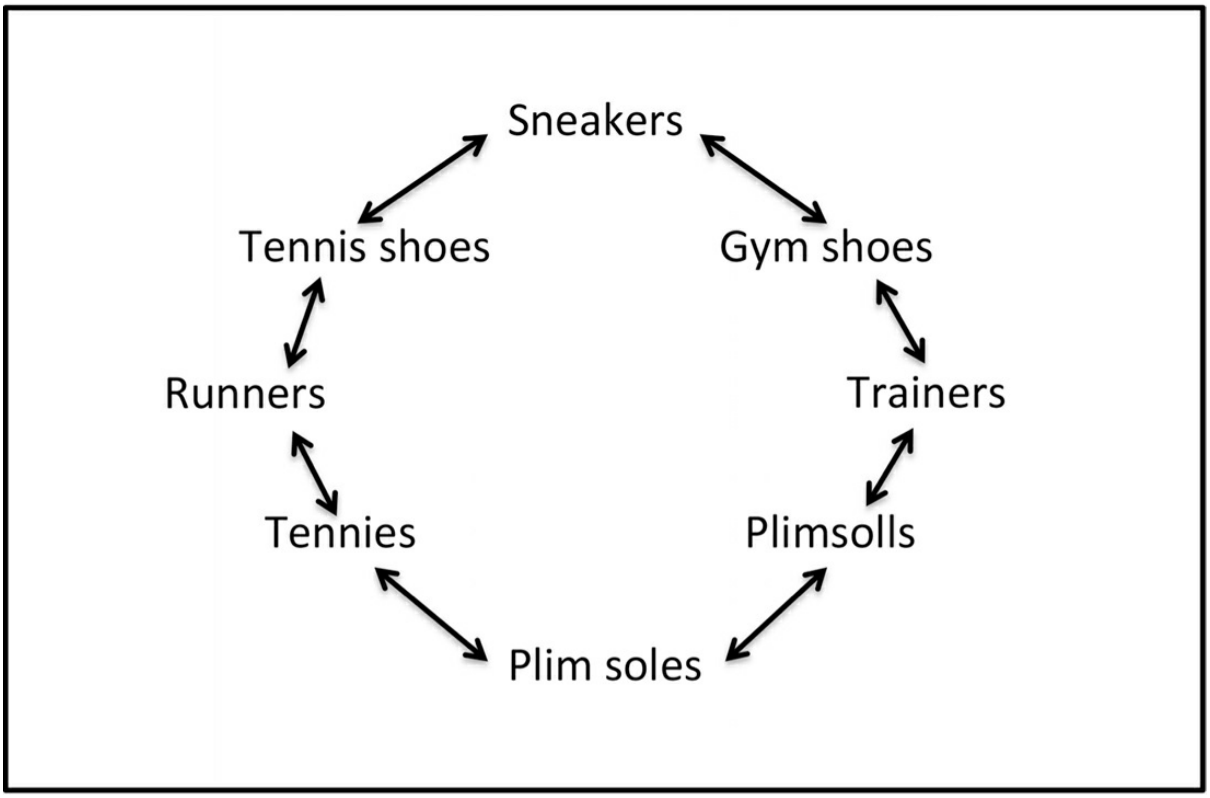


Figure 10.1. An Illustration of the *Synonym Ring* Concept.

Synonym Rings

According to Z39.19, a *synonym ring* is a different kind of controlled vocabulary. Its function is not so much to provide a list of terms to assign to documents during the indexing or cataloging process, instead it is used during retrieval activities. It is a behind-the-scenes instrument used in retrieval tools to connect equivalent terms (i.e., synonyms). A synonym ring is made up of a number of equivalence tables or lists. It is used specifically to broaden a search to enhance information retrieval, so that when a user searches one term within the cluster, all the terms can be searched. One term *may* be the authorized one if the terms are from a thesaurus, but authority control is not the focus of this type of controlled vocabulary—it is about improving retrieval among natural language variants. An illustration of the concept is provided in Figure 10.1.

Taxonomies

The third type of vocabulary, a *taxonomy*, is a hierarchically ordered list of terms. It is an orderly classification illustrating a defined knowledge domain. A taxonomy, though, often contains preferred terms only; synonymous terms are not listed. A taxonomy may differ from a thesaurus (the fourth type of vocabulary) in that it generally has shallower hierarchies and a less complicated structure. Often the term is used for a scientific list, such as Linnaeus's eighteenth-century biological taxonomy, but it does not have to be. A taxonomy can contain terms from any discipline. For example, one can create a simple taxonomy of metadata types:

Metadata
- **Administrative metadata**
  - **Preservation**
  - **Recordkeeping**
  - **Rights**
  - **Technical**
  - **Use**
- **Descriptive metadata**
  - **Analytical**
  - **Contextual**
- **Structural metadata**
  - **Behavioral**

Sometimes, the term *taxonomy* is used interchangeably with the term *classification*, or with the fourth vocabulary type: the *thesaurus*. Additional information about taxonomies is found in the next chapter.

Thesauri and Subject Heading Lists

According to Z39.19, *thesauri* are the most complex type of controlled vocabulary in use in the library and information science professions. A thesaurus is an authority-controlled list of terms in which semantic relationships are identified. Often the focus of a thesaurus is on hierarchical structures, but reciprocal equivalence and associative relationships are also included. Thesauri are usually made up of single terms and bound terms representing single concepts (often called *descriptors*). Bound terms occur when some single concepts can only be represented by two or more words. For example, the words *type, A,* and *personality* cannot be separated into discrete units without losing the meaning. The entire phrase, *Type A Personality*, is necessary for expressing the concept.

Since the mid-twentieth century, numerous thesauri have been created for a vast number of different subject areas. Some thesauri are publicly available online, but others are proprietary products of a particular institution or database and cannot be accessed without a subscription. Some thesauri comprise terms from multiple disciplines, such as the *UNESCO Thesaurus,*[5] which covers education, culture, natural sciences, social and human sciences, communication, and information, and the *American Folklore Society's Ethnographic Thesaurus (ET),* designed to provide access to resources about folklore, ethnomusicology, cultural anthropology, and other related fields.[6] Other thesauri focus primarily on narrower subject areas, such as the *NASA Thesaurus,*[7] covering the aerospace industry. Later in this chapter, some representative thesauri are described briefly. The following example comes from the *Thesaurus of ERIC Descriptors (ERIC)*,[8] which is used to index education materials:

Early Reading
Scope Note: Reading by children before they reach school age

Broader Terms: **Reading**

Narrower Terms: N/A
Related Terms: **Beginning Reading**
Early Experience
Emergent Literacy
Prereading Experience
Reading Readiness

A *subject heading list* is a specialized type of thesaurus, primarily created in libraries, that allows for more complexity in the structure of its authorized terms. For example, the LCSH equivalent subject heading to the ERIC term just given is: **Reading (Early childhood)**. In addition to single terms and bound concepts, subject heading lists tend to allow compound phrases and intricately constructed strings of terms. Subject heading strings may append one or more subdivisions (based on geographic, chronological, form, or additional topical aspects) to an initial term or phrase in order to create additional context for the concept. The following is an example from LCSH of a string composed of the name heading for the U.S. Army, with three topical subdivisions added to bring out more specific aspects of the topic.

United States. Army—Guided missile personnel—Training of—Aids and devices

Among the best known and most used subject heading lists are *Library of Congress Subject Headings* (LCSH), *Sears List of Subject Headings (Sears),* and *Medical Subject Headings* (MeSH). A brief description of each is provided later in this chapter, and more detailed descriptions can be found in *Introduction to Cataloging and Classification* by Daniel N. Joudrey, Arlene G. Taylor, and David P. Miller. Additional sources are also suggested at the end of this chapter.

Comparing Thesauri and Subject Heading Lists

As stated, subject heading lists have been created largely in library communities, while thesauri have been created largely in indexing communities. This does not mean, however, that libraries cannot or do not use thesauri or that an indexing service might not employ a subject heading list; it is simply a generalization about the origins of these tools. Both types attempt to provide subject access to information resources by providing terminology that can be consistent and reliable rather than uncontrolled and unpredictable. Both choose preferred terms and make references from non-used terms. Both provide structural hierarchies so that terms are presented in relation to their broader terms, narrower terms, and related terms.

Although Z39.19 does not distinguish subject heading lists from thesauri, there are some distinctions worth noting. Thesauri are more strictly hierarchical. Because they are usually made up of single terms, each term often has only one broader term. The rules in Z39.19[9] that have to do with identifying broader, narrower, and related terms are much easier to follow when working with a single-term system than when working with a system that includes phrases and compound headings. Thesauri are typically narrower in scope, comprising terms from one specific subject area. Subject heading lists tend to be more general in scope, covering a broad subject area or, indeed, the entire reach of knowledge. Thesauri are more likely to be multilingual than are subject heading lists. Again, because single terms are used, equivalent terms in other languages are easier to find and maintain.

Thesauri and subject heading lists are the most widespread forms of subject vocabularies used in libraries, archives, and museums. Discussions of some of the more frequently used ones are provided later in this chapter. Some additional forms and structures related to terms and concepts are also addressed (e.g., ontologies, folksonomies). First, however, we discuss some challenges and principles associated with vocabulary control.

CONTROLLED VOCABULARY CHALLENGES

In the process of creating a controlled vocabulary, there are certain difficulties that must be addressed. An understanding of these problems can enhance one's ability to use a particular existing vocabulary. The following are the issues addressed in this section:

- Specific versus General Terms
- Synonymous Concepts

- Word Form for One-Word Terms
- Sequence and Form for Multi-word Terms and Phrases
- Homographs and Homophones
- Qualification of Terms
- Abbreviations and Acronyms
- Popular versus Technical Terms
- Subdivision of Terms
- Compound Concepts

Examples are provided to illustrate these challenges.

Specific versus General Terms

The level of specificity must be decided at the outset of establishing a controlled vocabulary. Various lists may have different thresholds for how specific the terminology will be. The heading **Cats** is not as specific as **Longhair cats.**

Cats
  Longhair cats
    Persian cat
      Silver Persian cat

Longhair cats is not as specific as **Persian cat**, which in turn is not as specific as **Silver Persian cat**. Varying levels of semantic depth may be found in different vocabularies because they serve institutions of different sizes and kinds, different audiences, different age levels, and so on. In LCSH, for example, the most specific term available is **Persian cat**; in *Sears,* a smaller vocabulary usually serving less extensive collections, the most specific term is **Cats**, although an instruction is given that, if needed, a term for a specific breed of cat may be created.

To some extent the decision on this matter is based on the types of users who are expected to search for headings from the list and upon the nature of the information resources that are to be assigned terms from the list. If the collection has mainly general kinds of information, then **Cats** is probably sufficient, even to cover a few more specific items. If the users are children likely to be looking for general kinds of information, then again **Cats** is probably sufficient as the most specific level. A vocabulary used to describe a collection in a veterinary school, however, most likely will need to be more specific.

Synonymous Concepts

The English language rarely has absolutely true synonyms—that is, situations where two words mean the same thing and have no variations in nuance. However, there are multitudes of synonymous words and phrases that mean so close to the same thing that they can be interchanged for each other. Hans Wellish wrote, "Authors have, therefore, great freedom in the choice of terms and may use several words for the same concept, which may be admirable from the point of view of style, but would be disastrous when transferred unchanged to an index."[10] These are the terms that make keyword searching so problematic and frustrating.

In the creation of a controlled vocabulary it is necessary to identify all the synonymous and nearly synonymous terms that should be brought together under a single authorized term. For example, do the terms *apparel, clothes, clothing, costume*, *dress,* and *garments* all mean the same thing? If not exactly the same, or if they have different nuances, are the differences important enough to warrant separate vocabulary terms for them? Is there enough of a logical distinction among them that some information resources should be placed under the term *garments,* others under *clothing,* and still others under *clothes?* Would users (or indexers and catalogers) understand the distinctions among them? If not, then perhaps they should be treated as synonyms despite their slight differences. When making such a determination, one should take into account which of the terms is best known to the intended users; but with regional, national, and international differences in English-language usage, a decision may have to be arbitrary.

A key challenge in choosing among synonymous or nearly synonymous terms is ensuring that the terms chosen are neutral and inoffensive. Over the years various controlled vocabularies have been criticized for their choices of particular terms. Most often, this has focused on the ways that various groups of people—often marginalized groups of people—have been identified in the vocabularies. At times, terms and cross-references that are now considered racist, sexist, anti-Semitic, Islamophobic, homophobic, transphobic, or otherwise offensive have been included as acceptable terms in vocabularies. Much of the attention given to these problems stems from the writings of Sanford Berman,[11] Hope Olson,[12] and others who have been concerned about the representation of groups of people in controlled vocabularies and classification systems.[13] Although their work is somewhat beyond the scope of this book, readers interested in exploring these issues are encouraged to see Berman's seminal 1971 tract on LCSH, *Prejudices and Antipathies*, for an introduction to some of the problems.[14] In a review of Berman's work 30 years later, Stephen Knowlton found that nearly two-thirds of the problematic headings that Berman identified in *Prejudices and Antipathies* had been partially or fully revised to address Berman's concerns.[15] Despite this progress new issues continue to arise—most recently in a surprisingly widespread debate over the use of the term *illegal aliens*, which many consider pejorative and dehumanizing. Others, however, have argued for its continued usage—the debate over which has spilled into the chambers of the U.S. Congress.[16] At the time of this writing, the issue is not yet resolved. Although rare, this is *not* the first time a government body has involved itself in the details of cataloging. (See the discussion of Anthony Panizzi in Chapter 2.)

Word Form for One-Word Terms

Words in English often have more than one form that can mean the same thing (e.g., *clothing* and *clothes*). Also, language evolves, and as it does, a concept has a tendency to be expressed first as two words, then as a hyphenated word, then as one word (e.g., *meta data, meta-data, metadata*). Sometimes all three forms appear in use at the same time. British and American spellings give us another case of word form difference (e.g., *catalogue* and *catalog*). Prefixes to a word can create a word with a different meaning, but in some cases, where the meaning is opposite, it would not make sense to use both terms in the controlled vocabulary because one of the concepts can never be discussed without the other (e.g., *armament* and *disarmament*; *equality* and *inequality*).

A major word form difference is singular versus plural. There is no rule on which form to use; each vocabulary may approach it differently, following the policies of the creating agency. Most of the time the plural has the broadest coverage (e.g., *videocassettes* rather than *videocassette*); but at times the singular is broader (e.g., the term *apple* can apply to both the fruit and the tree, while *apples* refers only to the fruit). Sometimes the singular and the plural forms of a word have different meanings (e.g., in *Sears*, the heading **Art** refers specifically to visual art; the heading **Arts** refers to a broader concept that includes visual arts, literature, and the performing arts).

Sequence and Form for Multi-word Terms and Phrases

In some controlled vocabularies there are terms and phrases made up of two or more words. Some of these headings are modified nouns (e.g., **Environmental education**); others are phrases with conjunctions or prepositions (e.g., **Information theory in biology**); and a third group has qualifiers added in parentheses (see discussion below). A problem in constructing such terminology in a controlled way is being consistent in the order and form of the individual words used. For example, the phrases *energy conservation* and *conservation of energy resources* mean the same thing. The first phrase places the concept with other headings beginning with the word *energy*; the second phrase puts the concept with other topics having to do with conservation. If the list creates such phrases, it must be certain to have references from every possible construction of the phrase—referring from them to the construction that was chosen.

Some controlled vocabularies (notably LCSH) present certain multi-word terms and phrases in inverted order (e.g., **Education, Bilingual**; **Asylum, Right of**). Much of this was done in the past in order to collocate a group of headings on a broad concept with subconcepts arranged alphabetically below it. Thus, instead of *bilingual education* being found in the Bs and *higher education* being found in the Hs, both were found in the Es as **Education, Bilingual** and **Education, Higher**. Research has shown that few users think of such phrases in inverted order but instead look for them in direct order. In LCSH few new inverted forms are being established, but previously established ones still exist. One therefore finds inconsistencies such as **Medical education** and **Communist education** juxtaposed with **Education, Humanistic** and **Education, Greek**.

Homographs and Homophones

Homographs are words that look the same but have different meanings. *Mercury* can be a liquid metal, a planet, a car, or a Roman god; *bridge* can be a game, a structure spanning a chasm, a device connecting two computer networks, a location on a ship, or a dental device, among other meanings. In a controlled vocabulary there must be some way to differentiate among the various meanings. Two common ways are either to use qualifiers to distinguish between the terms or to choose synonyms as the preferred terms. In the world of art and architecture, the term *structures* can be used in more than one way; consequently, AAT has created two separate qualified headings to address the different meanings: **structures (single built works)** and **structures (structural elements)**.

Homographs may or may not be pronounced the same. For example, *mare* pronounced as one syllable with a silent *e* is a mature female horse; *mare* pronounced as two syllables is a large, dark area on the moon (derived from the Italian *mare,* meaning *sea*). Traditionally, different pronunciations did not play a role in creating a controlled vocabulary, because the vocabulary was treated visually.

Homophones, which are words that are spelled differently but pronounced the same, have also been ignored in controlled vocabularies in a visual world (e.g., *moat* and *mote; fowl* and *foul)*. However, because what appears on computer screens is now quite regularly read aloud electronically to people with visual impairments, we need to give attention to pronunciations of homographs and to distinguishing among homophones.

Qualification of Terms

As mentioned above, one of the ways of dealing with homographs is to add a *qualifier*. A qualifier is an addition or an annotation added to a term to clarify its meaning. For example, the homograph *bridges* is represented in LCSH in four ways to address some of its various meanings:

- **Bridges—**the unqualified heading is used to represent structures spanning open spaces or water
- **Bridges (Computer network)**
- **Bridges (Dentistry)**
- **Bridges (Graph theory)**

Qualifiers are also used to differentiate usages of a word in different settings. For example:

- **Adultery (Aztec law)**
- **Adultery (Byzantine law)**
- **Adultery (Canon law)**
- **Adultery (Germanic law)**
- **Adultery (Inca law)**
- **Adultery (Islamic law)**
- **Adultery (Jewish law)**
- **Adultery (Roman law)**
- **Adultery (Yanzi law)**

In addition, qualifiers can be used to help identify the context of unfamiliar words. For example, the name *Yanzi* is identified as **Yanzi (African people)** in the authorized heading for that group.

Abbreviations, Acronyms, and Initialisms

Abbreviations, acronyms, and initialisms can cause problems in adding concepts to controlled vocabularies. *Abbreviations* are shortened forms of words. *Acronyms* are abbreviations made up of initial letters of words from a phrase, and the resulting group of letters is pronounced as a word (e.g., *radar*). *Initialisms* are abbreviations made up of initial letters of a phrase, but each letter is pronounced separately (e.g., ABC for the

American Broadcasting Company). Traditionally, the determination whether to spell out abbreviations, acronyms, and initialisms was based on the intended users of the controlled vocabulary and their expected knowledge. With a move to more global retrieval, this practice has been reconsidered. Without the ability to assume a certain population, it is best to assume that they should be spelled out. A few, however, have global recognition. In the English version of the *UNESCO Thesaurus*, the acronym **AIDS** is the preferred term, but alternative forms—*Acquired Immunodeficiency Syndrome* and *HIV/AIDS*—are provided as cross-references. In the Spanish and French versions of the thesaurus, the acronym **SIDA** is the preferred form, with its own references.

Popular versus Technical Terms

When a concept can be represented by both technical and popular terminology, the creators of a controlled vocabulary, sometimes called *taxonomists* in specialized subject communities, must decide which will be used. For example, the National Library of Medicine's MeSH uses **Neoplasms**, whereas LCSH uses **Cancer**. If the list is intended to be used for information resources that will be retrieved by a specialized audience only, then specialized terminology is justified. Again, though, in a global information world, one can no longer be certain of a particular audience; for example, anyone with Internet access is able to view MeSH terminology, not just medical specialists. Perhaps this is another area where the kind of authority control discussed in Chapter 8 should be put into use. With international authority control, technical terms and their equivalent popular terms could reside on the same record, which could be activated by a search on any of the equivalent terms.

Subdivision of Terms

Subdivisions are used in controlled vocabularies that precoordi-nate terms. A *subdivision* is a term or phrase appended onto a subject heading to provide additional specificity. The following are some of the uses of subdivisions in LCSH:

- to separate by form (e.g., **Chemistry—Dictionaries**)
- to show treatment of only a part of the larger subject (e.g., **Merchant marine—Officers**)
- to show special aspects of the larger subject (e.g., **Merchant marine—Watch duty**)
- to show geographical or chronological limitations (e.g., **Architecture—Great Britain—19th century**)

Compound Concepts

As mentioned earlier, most thesauri primarily comprise single words and bound terms representing single distinct concepts. Other controlled vocabularies, however, include multitopic concepts in the form of phrases. When creating a controlled vocabulary, the taxonomist must determine how much compounding is desirable or acceptable, when to employ compounding (rather than employing single concepts), and what forms these phrases should take (discussed in the section on *Sequence and Form for Multi-word Terms and Phrases* above). The following are some examples of compound headings found in LCSH:

- **Animals, Prosecution and punishment of**
- **Comic strip characters in motion pictures**
- **Church work with cowgirls**
- **Diplomatic and consular service, Communist countries**
- **Marine invertebrates as pets**
- **Running races in rabbinical literature**

As you can see in the examples, multiple topics are being brought out in each heading. Multitopic subject terms can create problems in organizing a controlled vocabulary; as more concepts are incorporated into a phrase, the relationships to other concepts in the vocabulary can become less clear. For example, how is

Animals, Prosecution and punishment of related to **Animals** or to **Prosecution** or to **Punishment**? This is especially difficult to determine when one learns of the heading's broader term: **Superstition**.

PRE-COORDINATION VERSUS POST-COORDINATION

Controlled vocabulary terms can be assigned either in a pre-coordinated fashion (i.e., the indexer constructs subject strings with main terms followed by subdivisions), or in a post-coordinated fashion that requires the searcher of the system to combine the terms (i.e., search the two or more terms together). When terms are pre-coordinated in the controlled vocabulary or are pre-coordinated by the cataloger or indexer, some concepts, subconcepts, place names, time periods, and form concepts are put together in subject strings. Here are some examples of pre-coordinated strings found in LCSH:

- **Money—United States—History—Colonial period, ca. 1600–1775**
- **Gold mines and mining—Accidents—California**
- **Cultural property—Protection—Law and legislation—Criminal provisions**

This does not mean that all concepts needed to describe the aboutness of a particular resource will be placed in the same subject string. In most pre-coordinated systems in use today, there are several pre-coordinated strings per surrogate record, because controlled vocabulary instructions generally control the length of strings. Too many concepts in the same string can be confusing for both catalogers and users.

In post-coordinated systems, each concept is entered into a surrogate record discretely, without stringing together subconcepts, place names, and the like. Searchers, therefore, must combine terms in a search box using Boolean operators (e.g., AND, OR, NOT). Keyword searching is the ultimate form of post-coordinate indexing. The following represents concepts added to a record using a post-coordinated approach to subject data:

- **Money**
- **United States**
- **Colonial period, ca. 1600–1775**

This, however, shows only one topic for a straightforward resource. When several subjects are addressed in the same resource, it can be unclear how the separate concepts are connected. For example, the following strings could be created for a resource that compares gold mining in California in the nineteenth century to diamond mining in South Africa in the twentieth century:

- **Gold mines and mining—California—History—19th century**
- **Diamond mines and mining—South Africa—History—20th century**

The pre-coordinated strings make clear the places and times in which the major topics are addressed. With a post-coordinated approach, this is not true:

- **Diamond mines and mining**
- **Gold mines and mining**
- **California**
- **South Africa**
- **History**
- **19th century**
- **20th century**

It is not quite clear if this resource is about

- diamond mining in California,

- gold mining in the twentieth century,
- diamond mining in South Africa in the nineteenth century,
- gold mining in South Africa, or something else.

Pre-coordinated strings provide context.

With the use of most controlled vocabularies, the searcher must still do some post-coordination, even if the cataloger or indexer has already pre-coordinated some concepts. It is a matter of degree. LCSH, for example, has no pre-coordinated terminology for the concept *dancers and musicians.* It is up to the user who wants a work covering this concept to search the headings **Dancers** AND **Musicians** using Boolean logic and then to determine which metadata records have both terms in their controlled vocabulary sections.

Post-coordination is sometimes confused with depth indexing, but these are very different concepts. Although it is often the case that the numerous discrete terms in a record that result from depth indexing are used in a post-coordinated system, depth indexing does not require the indexer to assign only single terms. There is nothing to keep a depth indexer from stringing together concepts with subconcepts or modifiers to make more meaningful pre-coordinated index terms. It depends on local indexing policies or rules associated with the particular controlled vocabulary being used.

GENERAL PRINCIPLES FOR CREATING CONTROLLED VOCABULARIES

There are some general principles that apply to the creation of controlled vocabularies. These are to be distinguished from principles that come into play when applying particular vocabulary terms. The general principles discussed here for creating vocabularies are *specificity, literary warrant,* and *direct entry.*

Specificity

Specificity is the level of semantic depth found in a particular controlled vocabulary. For example, LCSH has greater specificity in its established subject headings than does *Sears.* This is evident from the greater hierarchical depth in the concepts that are found in LCSH and is made physically obvious by the relative size of the printed volumes. For example, the very specific **Canned raspberries** is found in LCSH; it is a narrower term found under **Canned berries**, which is a narrower term found under the broader heading **Canned fruit**, which in turn is a narrower term under **Canned foods**.

Food
    Canned foods
        Canned fruit
            Canned berries
                Canned raspberries

In *Sears,* the choices are

(1) **Canning and preserving**, which is used for canned foods in general,

(2) **Berries**, which is broader than the topic, or

(3) **Fruit—Preservation**, which seems to be more about the process of preserving fruit than a particular product.

The instructions in *Sears* say that specific types of foods with the subdivision**—Preservation** can be added as needed. This allows the vocabulary to be supplemented when more specificity is deemed necessary. Subject heading lists and thesauri created for specific subject fields or disciplines, such as MeSH for the field of medicine, are even more specific in their headings than LCSH.

Literary Warrant

Controlled vocabulary lists tend to be created using the principle of *literary warrant.* This means that terminology is added to a subject heading list or thesaurus only when a new concept shows up in the literature and therefore needs to have specific terminology established. This is often described as a bottom-up approach because existing resources are used to expand the controlled vocabulary. Usually, no attempt is made to add new terminology to a list until it is needed for use in metadata records; otherwise, taxonomists would be predicting the future or basing new additions on what should be there, at least theoretically. In recent years, top-down approaches to creating subject languages have fallen out of favor. For example, in ERIC there are headings for **Asian Americans** and **Librarians**, but it does not contain a heading representing *Asian American librarians*, and it will not do so, unless there is educational literature related to that specific group.

Direct Entry

Another principle of creating controlled vocabulary terms is that of *direct entry,* which states that a concept should be entered into the vocabulary using the term that names it rather than treating that concept as a subdivision of a broader concept. For example, in LCSH there is currently a preference for a modified term to express a concept (e.g., **Railroad stations**) over the use of a broader term subdivided by a narrower term (e.g., **Railroads—Stations**).

GENERAL PRINCIPLES FOR APPLYING CONTROLLED VOCABULARY TERMS

The first step in applying controlled vocabulary terms takes place just after the cataloger or indexer determines the aboutness of a resource. Once the nature of content is understood, and an aboutness statement has been written, choices must be made to determine which subject concepts are to be represented by controlled vocabulary terms in the metadata record for the resource. Not every concept in an aboutness statement can be translated into a term from the controlled vocabulary that has been chosen for use in a particular institution or situation. Although controlled vocabularies are useful in describing key concepts, places, events, time periods, and objects (often in the form of nouns and various types of noun phrases), they are not always helpful in translating actions and relationships among those concepts. The indexer, therefore, must be familiar enough with the controlled vocabulary to know which ideas can be translated and which cannot.

In the aboutness statement, the most important concepts are selected as targeted searches in the controlled vocabulary. This may entail circling key words in the statement or compiling a list of terms. Translating these words into authorized descriptors or subject headings, however, may be more difficult than one might initially expect. One may have to follow several paths established by cross-references in the vocabulary before finding the appropriate terms. In some cases, it may take a staggering effort to remain faithful to the aboutness of the resource, while following specific application rules in addition to following general principles of applying controlled vocabulary. For example, subdivisions that look appropriate for the resource may be restricted in their use, leaving the cataloger or indexer feeling thwarted by the complex rules that accompany some controlled vocabularies. In other cases, one may discover that there is simply no appropriate way to translate a particular concept into a vocabulary.

The general principles discussed below are *specific entry* and *coexten-sive entry.* The number of terms assigned and what to do when the concept is not in the particular vocabulary are also discussed. Specific rules for individual controlled vocabularies are not addressed.

Specific Entry and Coextensive Entry

The principle of *specific entry* states that an aboutness concept should be assigned the most specific term that is available in the controlled vocabulary. For example, an information resource about musicians should be entered under **Musicians**, not under **Entertainers**, or if it is about pianists, and that term is in the vocabulary, then the entry should be **Pianists**, not **Musicians**. However, if the vocabulary does not get as specific as **Pianists**, or does not allow creation of such terms for specific categories of musicians, then the principle of specific entry calls for using **Musicians** as the most specific entry from that vocabulary. Specific entry allows an experienced user to know when to stop searching for an appropriate controlled vocabulary term. One does not have to keep trying broader terms, unless no information is found under the most specific terms.

It should be noted that the concept of *specific entry* is not the same as the concept of *coextensive entry. Coextensivity* is the idea that the subject headings applied to an information resource will cover all, but no more than, the concepts or topics covered in that resource; it is about matching exactly the subject entries to the limits and scope of the aboutness of the information resource. In order to have coextensive entry using LCSH, for example, an information resource about crocheting potholders requires two specific entries: one for **Crocheting** and another for **Potholders**. There is no one specific heading to cover these two concepts together. In order to have just one heading that is coextensive with the subject of such a resource, the heading would have to be *Crocheting potholders,* or *Potholders—Crocheting,* but these are not used in LCSH.

It once was true that *specific entry* could be treated in a relative way. In a small collection there might be only one or two items about musicians, although there might be several items about entertainers that include musicians as one type of entertainer; for that collection, there might be a decision that **Entertainers** would be the most specific heading used. Now that we are all essentially contributing to a global union catalog, however, we must follow the principle of specific entry in order for searching to be effective.[17] On the other hand, some have called for including broader terms in catalog records to assist less-experienced catalog users with navigating the system and finding the resources that they need[18] (e.g., a record could include headings for both **Persian cats** *and* **Cats**, even though the item focuses primarily on Persian cats). Or, if we could break out of tradition and make references in our retrieval tools from specific subject headings to broader ones, then "too specific" would not have to be a concern.

Number of Terms Assigned

There should be no arbitrary limit on the number of terms or descriptors assigned. If the conceptual analysis has been done at the summarization level, then the number of terms given should be the number that is needed to express that summary. Likewise, if the conceptual analysis has been performed in depth, the number of terms necessary to cover all of the concepts should be allowed. This, however, is not always the case. Some information institutions have local policies that restrict the number of subject headings assigned to an information resource. While this *might* help to cut down on the time spent indexing, it does the users no favors. For example, if the local policy states, "No more than three descriptors are to be added to a record," but the information resource is about *four* specific, discrete concepts, which one is to be left out of the record? The time spent deciding what topics to omit might in fact be lengthier than the time spent adding a fourth subject term.

Concepts Not in the Controlled Vocabulary

If a concept is not present in the controlled vocabulary, it should be represented temporarily by a more general concept, rather than simply adding unauthorized terms to the record (e.g., use **Artificial intelligence** until a new heading for *machine discovery* is established). The new concept should be proposed as a new addition to the subject list or thesaurus. And once (or *if*) the new concept is added to the vocabulary, a change in the metadata record should be made. This means that the cataloger or indexer should flag the record for updating; otherwise, it might be forgotten or lost among a sea of records. It should be noted, however, that some controlled vocabulary systems allow the cataloger or indexer to add their own terms as needed.

INDEX TERMS FOR NAMES

Although the terms for most topical concepts are handled by controlled subject vocabularies, a separate authority file generally controls most proper names. For example, the Library of Congress (LC) maintains the LC/NACO Authority File (LCNAF) for names of persons, corporate bodies, geographic names of political entities, titles of works, and name/title combinations.[19] Geographic names that are *not* political entities and other names such as names of archaeological sites are also controlled by LC, but as subject headings rather than as names. The art world also has several name authority files maintained by the Getty: the *Union List of Artist Names* (ULAN), the *Cultural Objects Name Authority* (CONA), and the *Getty Thesaurus of Geographic Names* (TGN).[20]

MECHANICS OF CONTROLLED VOCABULARIES

A key feature of thesauri and subject heading lists is that they identify relationships among terms and make connections (i.e., references) between related concepts. The most commonly identified relationships among terms are organized into three primary types: (1) *equivalence relationships*, (2) *hierarchical relationships*, and (3) *associative relationships*. Each is discussed below.

Equivalence Relationships

Thesauri (including many subject heading lists) operate by choosing a preferred way of expressing a concept and then making certain that synonymous ways of expressing the concept will be connected to the preferred terminology. This practice identifies and addresses *equivalence relationships*. Traditionally, the unused (or unauthorized) terminology appears in the vocabulary listed under the preferred terminology and is often preceded by the abbreviation **UF** meaning *used for*. The unauthorized terms also appear in the list as *entry vocabulary* to act as pointers to the chosen terms (e.g., as a **USE** reference). For example, under the heading **House plants**, two unauthorized terms are listed.

House plants
UF Houseplants
Indoor plants

Elsewhere in the vocabulary, two reciprocal references are made at the place in the list where the unauthorized terms appear. For example:

Houseplants
USE **House plants**
. . .
Indoor plants
USE **House plants**

Equivalence relationships not only address truly synonymous terms, such as *clothing* and *clothes*, but also *nearly synonymous* terms. These are terms that are not exactly equivalents, but they are close enough that it is deemed impractical to establish both terms in a controlled vocabulary. For example, *seawater* and *ocean water* may not be precisely identical (obviously the source of the water differs), but is there enough of a difference that some users would only want resources about ocean water but not about seawater or vice versa? In LCSH, **Seawater** was chosen as an authorized term and *Ocean water* became a reference.

Also included under the category of equivalence relationships are lexical variants and some antonyms, both mentioned above. With lexical variants, we want to ensure that one spelling is chosen and alternative spellings are references.

Catalogues
USE **Catalogs**

With some (but not all) terms that have opposite meanings, we may establish one and refer from another if the terms are so interconnected that one cannot be understood without the other.

Inequality
USE **Equality**

Hierarchical Relationships

Thesauri also keep track of the hierarchical relationships of a concept. In most, these are designated as *broader term* (**BT**) and *narrower term* (**NT**) relationships, although some thesauri, such as AAT, may not use the traditional relationship indicators (relying on hierarchical displays instead). The relationship between **BT** and **NT** is similar to a parent–child relationship. The broader term is the parent; it has less specificity. The narrower term is the child term; it is more specific, more focused. The relationship is reciprocal. So if a **BT** points to an **NT**, then the entry for the **NT** must point to the **BT** in the thesaurus (*note*: references from narrower terms to broader terms typically do not appear in most catalogs).

There are different kinds of hierarchical relationships that may be designated as simply broader and narrower terms in a general vocabulary but which may be designated more specifically in some thesauri. These include

- **Genus–Species** (or Class–Class member) relationships: This type of relationship indicates that the narrower terms are a *type of* or *kind of* whatever the broader term represents. In the following example, we can see that all of the **NT**s are a type of **Building**.

  Buildings
  NT Auditoriums
  Church buildings
  Clubhouses
  Garages

- **Whole–Part** relationships: The name of this type of relationship is rather self-evident. The **BT** would represent the whole, while the **NT**s would represent parts. In the example, the **NT**s are parts of a head.

  Head
  NT Brain
  Ear
  Face
  Hair
  Mouth
  Nose

- **Instance** (or Generic topic–Named example) relationships: In this relationship the **BT** is a particular category of thing, and each **NT** is an example or an instance of the thing. For example, the **Aegean Sea** is not a type or kind of sea, it is an instance of a sea.

  Seas
  NT Adriatic Sea
  Aegean Sea
  Arabian Sea
  Baltic Sea
  Caribbean Sea

Associative Relationships

The final type of relationship highlighted in many thesauri is the *associative relationship*. These are usually designated as *related term* or **RT** relationships. These, too, are reciprocal. These are the least well defined of the relationships. They indicate a relationship that might be of interest, which is not a hierarchical or equivalence-based relationship. Some controlled vocabularies do not enumerate specific types of RT relationships to include, whereas others might provide strict guidelines. There are different kinds of related-term relationships. A few include

- one term needed in the definition of the other (e.g., *stamps* is needed in the definition of *philately*);
- meanings of two terms overlap, or two terms may be used interchangeably, yet are not synonyms (e.g., *carpets* and *rugs*); and
- linking of persons and their fields of endeavor (e.g., *attorneys* and *law*).

In the following full example of an entry in a subject heading list, the subject heading (i.e., the preferred term in boldface type) is shown in relationship to its **UF**s, **BT**, **NTs**, and **RTs**. In addition, a *see also note* (**SA**) provides additional information about relationships with other headings of interest.

Maintenance

UF	Preventive maintenance
	Upkeep
BT	Maintainability (Engineering)
RT	Repairing
	Service life (Engineering)
SA	subdivision **Maintenance and repair** under kinds of objects, including machinery, vehicles, structures, etc., e.g., **Automobiles—Maintenance and repair**; **Dwellings—Maintenance and repair**; **Nuclear reactors—Maintenance and repair**
NT	Buildings—Repair and reconstruction
	Grounds maintenance
	Military bases—Maintenance
	Plant maintenance

As stated, all the relationships are reciprocal: under **Maintainability (Engineering)**, the narrower term **Maintenance** would be listed. Under **Repairing**, the related term **Maintenance** would be listed, and under **Grounds maintenance**, the broader term **Maintenance** would be listed.

Lexical Relationships

Relationships can also be expressed in other ways in some types of vocabularies. For example, the following are relationships shown in the general lexical database WordNet:[21]

- **Synonyms**: Terms that have the same, or nearly the same, meaning and often can be substituted for each other. A synonym is like the **UF** relationships and Use references in traditional thesauri.
- **Coordinate terms**: Terms that might be called siblings; they all have the same parent term. A coordinate term is *somewhat* similar to the **RT** relationships in traditional thesauri, although related terms in the same hierarchy are not articulated in many thesauri.
- **Hypernyms**: These are parent terms. It is a category comprising all of the instances that are "kinds of" the hypernym (e.g., *family* is a hypernym for *nuclear family, extended family, foster home*). A hypernym is like some **BT** relationships in traditional thesauri.
- **Hyponyms:** These designate child terms. A hyponym is a member of a class (e.g., *nuclear family* is a hyponym of the class *family*). A hyponym is like some **NT** relationships in traditional thesauri.
- **Holonyms**: A holonym is the name of the whole of which the meronyms are the parts (e.g., a *family* has as its members: *children, parents, sister, siblings*). A holonym is like some **BT** relationships in traditional thesauri.
- **Meronyms:** These terms designate constituent parts or members of the whole (e.g., *sister* is a meronym of *family*). A meronym is like some **NT** relationships in traditional thesauri.
- **Antonyms**: Terms that have opposite meanings (e.g., *hot* is an antonym of *cold*). Antonyms are not addressed in most traditional thesauri.

Most controlled vocabularies display terms in alphabetical order with all of their relationships enumerated below the entry, as seen in the **Maintenance** example, but some also provide a hierarchical display of the vocabulary terms (e.g., AAT). In the latter, the terms are placed in juxtaposition to each other so that one can visualize broader and narrower relationships. Such lists are helpful in seeing where a term fits within an entire hierarchy, not just its relationship to the terms immediately above and below it.

CONTROLLED VOCABULARY STANDARDS

In this section, several thesauri and subject heading lists commonly used in libraries, archives, and museums are described. These include the following:

- *Library of Congress Subject Headings* (LCSH)

- *Sears List of Subject Headings* (*Sears*)
- *Medical Subject Headings* (MeSH)
- *Library of Congress Genre/Form Terms for Library and Archival Materials* (LCGFT)
- *Library of Congress Demographic Group Terms* (LCDGT)
- *Art & Architecture Thesaurus* (AAT)
- *Thesaurus of ERIC Descriptors* (ERIC)
- *Faceted Application of Subject Terminology* (FAST)

Library of Congress Subject Headings (LCSH)

Library of Congress Subject Headings (LCSH) is a list of authority-controlled terms that are used to represent the subject matter of resources. It is an all-purpose, multidisciplinary vocabulary; it is not restricted to a particular knowledge domain. LCSH comprises subject headings that have been established since 1898; it also contains references, scope notes, and subdivisions that may be used to make headings more specific. LCSH is the most comprehensive list of English-language subject headings in existence. It attempts to cover the world of knowledge and is the only subject heading list accepted as a worldwide standard.[22] It is used in libraries, archives, and cultural heritage institutions in many different kinds of settings, including countries other than the United States. “LCSH has been translated into many languages and is used around the world by libraries large and small.”[23]

LCSH is available in several formats. The longest-running format was the print version. For many years it was published in new editions every five years or so, but starting in 1988 with the 11th edition, new print versions of LCSH were typically produced once a year, most recently as a six-volume set. In 2013, however, LC changed its approach to providing access to cataloging documentation, announcing that it would no longer publish physical copies of the list. LCSH is still produced but only in electronic formats. A fully updated electronic version is available (1) by subscription through *Classification Web*,[24] (2) in the subject authority files accessible through OCLC,[25] and (3) at the LC Authorities website.[26] Each year a new edition is also produced for distribution as print-ready PDFs on the LC website.[27] LCSH is, however, updated continuously. Each month new headings and references are added, existing headings and references are revised, and, occasionally, a few headings and references are removed.

In the subject cataloging process, LCSH is used in conjunction with the *Subject Headings Manual* (SHM).[28] The SHM is an essential tool for anyone who wishes to apply subject headings appropriately. It contains approximately 300 numbered *instruction sheets* all prefixed with the letter *H* (e.g., H 1095, H 180). This manual gives the subject heading policies and practices employed at LC, which is responsible for the maintenance of LCSH. It was formerly issued in print form but is now freely available in PDF at LC’s website.[29] The SHM is also available through LC’s web-based documentation service, *Cataloger’s Desktop*, by subscription.[30] A sample from LCSH is shown in Figure 10.2. MARC-encoded authority records for two of the headings in Figure 10.2 are shown in Figure 10.3 (as displayed in OCLC) and Figure 10.4 (as displayed in *Classification Web*). *Note*: MARC records for subject terms do not contain narrower terms, because the narrower terms have their own MARC records containing the reciprocal broader terms.

Fascism (May Subd Geog)

[D726.5 (World history)]

[JC481 (Political theory)]

Here are entered general works on fascism including works on post-World War II fascist movements. Works on German fascism during the Nazi regime are entered under National socialism.

UF	Neo-fascism
BT	Authoritarianism
	Collectivism
RT	Corporate state
	National socialism
	Synarchism
	Totalitarianism
NT	Anti-fascist movements
	Fascist economics
	Fascist propaganda
	Labor unions and fascism
	Neo-Nazism

Fascism—Argentina

NT	Peronism
...	

Fascism—Posters

BT	Political Posters

Fascism and architecture (May Subd Geog)

UF	Architecture and fascism
BT	Architecture
...	

Fascism and youth (May Subd Geog)

UF	Youth and fascism
BT	Youth

Fascism in art (Not Subd Geog)

...

Fascist propaganda (May Subd Geog)

UF	Propaganda, Fascist [Former heading]
BT	Fascism
	Propaganda, Italian

Fascists (May Subd Geog)

NT	Jewish fascists
	Neo-Nazis
	Women fascists

Figure 10.2. Sample Entries from *Library of Congress Subject Headings* (LCSH) as Displayed in *Classification Web* (January 20, 2017). Available by Subscription at: http://classweb.loc.gov.

Authority File: Fascism

ARN 2152577

Rec stat c Entered 19860211 Replaced 19990331151728.9

Type	z	Upd status	a	Enc lvl	n	Source	
Roman	|	Ref status	b	Mod rec		Name use	b
Govt agn	|	Auth status	a	Subj	a	Subj use	a
Series	n	Auth/ref	a	Geo subd	i	Ser use	b
Ser num	n	Name	n	Subdiv tp	n	Rules	n

010 sh 85047355
040 DLC $b eng $c DLC $d DLC
053 0 D726.5 $c World history
053 0 JC481 $c Political theory
150 Fascism
450 Neo-fascism
550 Authoritarianism $w g
550 Collectivism $w g
550 Corporate state
550 National socialism
550 Synarchism
550 Totalitarianism
680 $i Here are entered general works on fascism including works on post-World War II fascist movements. Works on German fascism during the Nazi regime are entered under $a National socialism.
681 $i Note under $a National socialism

Figure 10.3. A Sample Authority Record for a Heading Shown in Figure 10.2. (Source: OCLC Connexion, Authorities—Record Number 2152577)

Sears List of Subject Headings (Sears)

Sears is still published in print form with a new single-volume edition coming out every four years or so.[31] The 21st edition was published in 2014, and the 22nd edition is expected in 2018. In addition to the print volume of *Sears*, there is an electronic version of the list available. Since 2011 EBSCO Information Services has made the *Sears* list available through their EBSCOhost database service.[32] The same basic information is provided in the online version, but the formatting of the entries is somewhat different.

Subject Record [Fascist propaganda]

ID: sh 85107461	Entered: 860211	Replaced: 20001016
008/06 Geo Subd: i-Indirect	008/11 SH System: a-LCSH	008/29 Ref Eval: a-Eval
008/07 Roman: |-No attempt	008/15 Subj Use: a-Appropriate	008/31 Rec Upd: a-Can be used
008/09 Kind Rec: a-Estab hdg	008/17 Type Subd: n-Not applic	008/33 Level Estab: a-Fully

010	$a sh 85107461
035	$a (DLC)sh 85107461
035	$a (DLC)4757046
035	$a (DLC)sp 85107461
035	$a (DLC)262273
040	$a DLC $c DLC $d DHMM
150	$a Fascist propaganda
450	$w nne $a Propaganda, Fascist
550	$w g $a Fascism
550	$w g $a Propaganda, Italian

Figure 10.4. A Sample Authority Record for a Heading Shown in Figure 10.2, as Displayed in *Classification Web*. Available by Subscription at: http://classweb.loc.gov. (Source: *Classification Web*—Record Number 229739)

The *Sears* list is intended for small collections used by persons with general information needs. Its main users are school and small to medium-sized public libraries. For most of its existence *Sears* has followed the lead of LCSH in format and in terminology choices. In its use of LCSH terminology, though, *Sears* has used only the more general terms and has not included the more specific terms or the ones geared for research audiences. In addition, *Sears* has fewer subdivisions.

In the last few editions, *Sears* has taken the initiative to make changes that were later followed by LCSH. For example, the change of **Afro-Americans** to **African Americans** was made first by *Sears.* In some cases, however, Sears has made changes that are not in alignment with LCSH. In 2002 the 17th edition of *Sears* replaced the heading **Indians of North America** with **Native Americans**; LCSH has yet to make a change of this nature. In 2016 *Sears* replaced the term *Illegal aliens* with the new heading **Unauthorized immigrants**. (At the time of this writing LC has not yet decided how the problematic heading, **Illegal aliens**, will be addressed in LCSH; it *may* be replaced with two separate terms: **Noncitizens** and **Unauthorized immigration**.) In 2008 a Spanish-language translation, *Sears Lista de Encabezamientos de Materia,* was published.[33] A sample from *Sears* is shown as Figure 10.5.

Fascism (May subdiv. geog.) **320.53; 321.9; 335.6**

Use for general materials on fascism. Materials on fascism in Germany during the Nazi regime are entered under **National socialism**.

UF Authoritarianism
Neo-fascism
BT **Totalitarianism**
NT **National socialism**
Neo-Nazis

Fascism—Canada **320.53; 971.06**

Fascism—United States (May subdiv. geog.) **320.53; 973.9**

Fashion (May subdiv. geog.) **391**

Use for materials on **. . .**

Figure 10.5. A Representation of Subject Heading Entries from the 21st Edition of the *Sears List of Subject Headings*.

Medical Subject Headings (MeSH)

MeSH is the controlled vocabulary created and maintained by the National Library of Medicine (NLM).[34] There is a listing for each term or concept, as in other controlled vocabularies, giving its number, a scope note, related terms, and so on. It also provides a strict hierarchical structure (referred to as a *tree structure*). Although it does not specifically identify terms as BTs or NTs, the tree structures indicate the hierarchical relationships among the terms. For each single term and phrase found in the main heading list, MeSH provides a tailored list of allowable qualifiers (i.e., subdivisions).

MeSH was first produced as a print product beginning in 1960 and became available in electronic form in 1975. Since 2008, MeSH has been available only as an online resource in the MeSH Browser.[35] The MeSH Browser is provided online, free for all those who wish to use it.

MeSH, which is updated annually, is used for providing subject access points on every bibliographic record created at NLM, whether it is for the MEDLINE database or the NLM catalog. It can be searched through the MeSH Browser or can be downloaded in XML, MARC, or ASCII formats. A sample from MeSH, along with its allowable qualifiers, is shown in Figure 10.6. The three places in MeSH's hierarchical tree structure that contain the entry are shown in Figure 10.7.

Figure 10.6. A Representation of a MeSH Entry and Its Allowable Qualifiers.

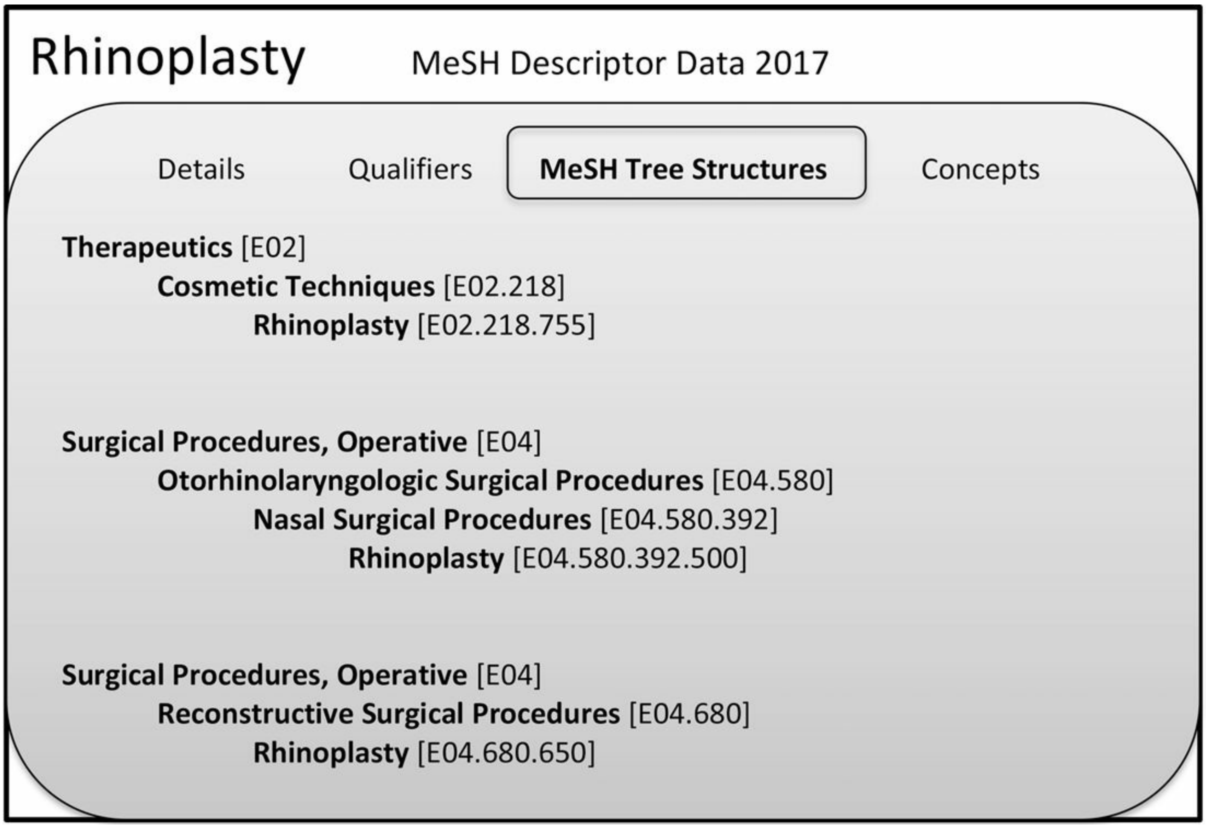


Figure 10.7. A Representation of the Tree Structure in MeSH.

Library of Congress Genre/Form Terms for Library and Archival Materials (LCGFT)

A relatively recent thesaurus, the *Library of Congress Genre/Form Terms for Library and Archival Materials* (LCGFT), is increasingly being used by libraries and archives to describe what a resource *is*, rather than what it is *about*. LCGFT contains *genre/form headings*, which generally represent the literary genre, artistic form, or publication format of a resource.

> In 2007 the Library of Congress [Policy and Standards Division (PSD)] began a project to develop genre/form terms. . . . Genres and forms may be broadly defined as categories of resources that share known conventions. More specifically, genre/form terms may describe the purpose, structure, content, and/or themes of resources.[36]

It does not include terms associated with ethnicity, nationality, audience, creation dates, geography, or popularity. For example, **Poetry** is included in LCGFT, but *French poetry* and *Medieval poetry* are not.

In 2011 LCGFT was published as a standalone vocabulary that may be used with LCSH, other subject access systems, or with descriptive cataloging content standards to provide values for form elements. It was developed through a series of special projects, some coordinated with outside groups to distribute the workload and to receive outside expertise in specialized areas (e.g., PSD worked with the American Association of Law Libraries' Classification and Subject Cataloging Policy Advisory Working Group on the terms for law materials). Currently, LCGFT contains genre and form terms from eight subject areas:

Topical Area	Examples
Cartographic materials	Atlases, Gazetteers, Maps
General materials	Biographies, Dictionaries, Menus, Wall charts
Law materials	Casebooks, Court rules, Statutes and codes
Literature	Black humor, Doggerel, Epic fiction, Puppet plays
Moving images	Film remakes, Road films, Television game shows
Music	Ballad operas, Gregorian chants, Peyote songs, Songbooks
Radio programs and nonmusical sound recordings	Field recordings, Radio adaptations, Western radio programs

Religious materials	**Church orders, Creeds, Myths, Sacred works**

At the time of this writing, a project with the Art Libraries Society of North America to develop art terms is in progress.

Unlike LCSH, LCGFT does not allow for subdivision of terms. It does, however, use many of the same formatting techniques and relationship indicators found in LCSH. It is likely that LCGFT will seem familiar to those who have used LCSH previously.

Historical drama
Plays set during a recognizable time period prior to the time in which they were written.
UF Chronicle histories (Drama)
Docudramas (Literature)
Documentary drama (Literature)
Verbatim drama
BT Drama
NT Pageants
Ritterdrama

Library of Congress Demographic Group Terms (LCDGT)

After the development of LCGFT, the Library of Congress PSD began work on a thesaurus of demographic terms. While developing LCGFT, it became clear that many terms used to describe the form or genre of creative works, such as **Children's poetry** and **Swahili drama**, often contained concepts that were related to creator or audience characteristics. In order to create a structurally pure thesaurus that does not include outside concepts, LCGFT excluded all demographic characteristics from its list of terms, and PSD began work on developing *Library of Congress Demographic Group Terms* (LCDGT) in 2013. The first edition of LCDGT was published in 2015.

A demographic group is any subset of the general population centered on a particular defined characteristic. LCDGT includes 11 categories of terms:

Category	Examples
Age groups	**Middle-aged people, Older people, Preteens**
Educational levels	**College juniors, Eighth grade students, Doctoral students**
Ethnic/Cultural group	**Basques (European people), Cree (North American people), Hispanic Americans**
Gender groups	**Boys, Intersex people, Females, Transgender people**
Language groups	**Central Khmer speakers, Marathi speakers, Spanish speakers**
Medical, psychological, and disability category	**Depressed people, Dyslexics, Leprosy patients**
National/Regional groups	**Asians, Bay Staters, Bhutanese, Germans**
Occupations and Fields of Activity	**Book editors, Catalogers, Lighting designers, University and college faculty members**
Religion groups	**Ashkenazim, Benedictines, Sufis**
Sexual orientation groups	**Bisexuals, Gay men, Lesbians**
Social groups	**Children of depressed people, Ecology students, Socialists**

Some terms may fit more than one category. For example, **Ashkenazim** belongs in both the *religion* and *ethnic* groups; **Boys** belongs to both the *gender* and *age* groups. Demographic terms may be used to represent audience or creator/contributor characteristics of a resource. The following is an example from LCDGT; it, too, is similar in structure to LCSH.

Cistercians
UF Bernardines
Cistercians of Common Observance members
White Monks
BT Benedictines
NT Trappists

Art & Architecture Thesaurus (AAT)

The *Art & Architecture Thesaurus* is intended to assist with verbal access to all kinds of cultural heritage information. Terms are provided for describing objects, textual materials, images, architecture, and material culture.[37] AAT is widely used in several communities: archives, libraries, museums, visual resources collections, and conservation agencies. It is arranged into eight facets (categories) that progress from the abstract to the concrete:

- **Associated Concepts**: abstract concepts and phenomena; theoretical and critical concerns, ideologies, attitudes, and social or cultural movements
- **Physical Attributes**: observable/measurable characteristics of artifacts and materials
    - Attributes and Properties
    - Conditions and Effects
    - Design Elements
    - Color
- **Styles and Periods**: terms for styles, periods, cultures, nationalities
- **Agents**: designations of people, animals, and plants
    - People
    - Organizations
    - Living Organisms
- **Activities**: actions, endeavors, and occurrences
    - Disciplines
    - Functions
    - Events
    - Physical and Mental Activities
    - Processes and Techniques
- **Materials**: substances used to produce structures or artifacts
- **Objects**: things
    - Built Environments
    - Components
    - Furnishings and Equipment
    - Object Genres
    - Object Groupings and Systems
    - Visual and Verbal Communication
- **Brand names**: objects, materials, and activities known by brand names

The facets are divided into one or more hierarchies. For example, one can see that the Activities facet is broken down into five hierarchies, each of which may be broken down further into additional conceptual hierarchies or clusters of terms organized into tree structures of broader and narrower terms.

AAT is available as a searchable web resource that provides full access to the preferred terms, alterative terms, and sources.[38] It can also be freely downloaded in two formats for incorporation into local retrieval tools, either as XML encoded records or as relational tables. The AAT dataset is also freely available in several different linked data formats (e.g., RDF, Turtle). The electronic version is updated continually. Extensive online documentation is available to help the novice searcher. An example of the type of data found in an AAT record is shown in Figure 10.8. On the AAT website, terms also display additional scope notes, as well as extensive lists of sources and contributors. To get a better sense of the full display, see the record for "still life painters" in the *Art & Architecture Thesaurus Online*.[39]

Click the icon to view the hierarchy.

ID: 300266118 **Record Type:** concept

still life painters (still life artists, <artists by subject or style of work>, ... People (hierarchy name))

Note: Artists who specialize in painting inanimate objects such as bowls of fruit, bottles, and flowers, which are often arranged in an aesthetic composition.

Terms:

still life painters (**preferred**,C,U,English-P,D,U,PN)
still life painter (C,U,English,AD,U,SN)
still-life painter (C,U,English,AD,U,N)
still-life painters (C,U,English,UF,U,N)
stillevenschilders (C,U,Dutch-P,D,U,U)
pintores de naturaleza muerta (C,U,Spanish-P,D,U,PN)
pintor de naturaleza muerta (C,U,Spanish,AD,U,SN)

Facet/Hierarchy Code: H.HG

Hierarchical Position:

Agents Facet
.... People (hierarchy name) (G)
.... people (agents) (G)
.... <people by occupation> (G)
.... <people in the humanities> (G)
.... <people in the arts and related occupations> (G)
.... <people in the arts> (G)
.... <people in the visual arts and related occupations> (G)
.... <people in the visual arts> (G)
.... artists (visual artists) (G)
.... <artists by subject or style of work> (G)
.... still life artists (G)
.... still life painters (G)

Additional Parents:

Agents Facet
.... People (hierarchy name) (G)
.... people (agents) (G)
.... <people by occupation> (G)
.... <people in the humanities> (G)
.... <people in the arts and related occupations> (G)
.... <people in the arts> (G)
.... <people in the visual arts and related occupations> (G)
.... <people in the visual arts> (G)
.... artists (visual artists) (G)
.... <artists by medium or work type> (G)
.... painters (artists) (G)
.... <painters by subject of work> (G)
.... still life painters (G)

Additional Notes:

...

Sources and Contributors:

...

Figure 10.8. A Partial Record from the Art & Architecture Thesaurus Online. (Courtesy J. Paul Getty Trust)

Thesaurus of ERIC Descriptors

ERIC is the acronym for the *Educational Resources Information Center*, which is a national information system designed to provide access to a large body of education-related literature. Among the documents that ERIC indexes besides journal articles are descriptions and evaluations of programs, book reviews, research reports, curriculum and teaching guides, instructional materials, position papers, computer files, and resource materials. These materials are indexed using terms from the *Thesaurus of ERIC Descriptors.*[40]

The print version of the thesaurus, last published in 2001, consists of four parts: the main alphabetical display, the rotated display, the hierarchical display, and the descriptor group display. The alphabetical display is similar to the display in LCSH (i.e., all of the terms in one alphabetical sequence). The rotated display provides an alphabetical index to every word in the thesaurus, including access to unused terms as well as main terms. The hierarchical display shows broader and narrower terms in their relationships to each other. The descriptor group display offers a kind of table of contents by placing all descriptors into a set of broad categories.

The most current version of the thesaurus is available as an online product. Its interface allows both searching and browsing the thesaurus.[41] It displays the ERIC descriptors in alphabetical order as search results and in the browse interface. Within each record for each term, broader and narrower terms are hyperlinked, allowing users to explore the hierarchical relationships among terms. Conveniently, the interface also allows users to search for a descriptor in the ERIC index directly from that term's record via a hyperlink marked "Search collection using this descriptor." A representation of an entry from the online display of the ERIC thesaurus is shown below.

Metadata
Scope note: Information that characterizes data, or the individual elements that describe and are used to provide access to an object, most often an information resource
Category: Information/Communications Systems

Broader Terms:	**Data**
Narrower Terms:	N/A
Use this term instead of:	**Metainformation**
Related Terms:	**Cataloging**
	Indexing
	Information Retrieval

Faceted Application of Subject Terminology (FAST)

At the turn of the century, some catalogers began to express dissatisfaction with LCSH and other more complicated vocabularies. Many felt that there was less need for complex subject heading strings, especially when cataloging web resources and digital objects. Their desire was for a controlled vocabulary that retained the rich semantics of LCSH but was, at the same time, intuitive and easy enough to apply without having to have a master's degree.[42] A research group at OCLC, working with Lois Mai Chan of the University of Kentucky and Lynn El-Hoshy of LC's Cataloging Policy and Support Office (now PSD), explored the issue and decided to develop a new schema derived from LCSH, but specifically designed for applications that go beyond the traditional library catalog. Thus FAST (Faceted Application of Subject Terminology) was born.

FAST[43] is a faceted vocabulary, meaning that the terms are divided into defined categories representing particular aspects or angles of the subject matter. For example, some terms are topical, some are chronological, and so on. This means that in FAST, the LCSH string **Microsoft Corporation—Employees—Biography** is converted to three separate headings, each representing a separate facet. FAST has eight facets:

Category	Description	Examples
Topical	General subject matter	Cataloging
Geographic	Place names	Europe; Uruguay
Chronological	Time periods	2001–2009; 1984
Events	Occurrences or historical incidents	Cannes Film Festival; War of 1812
Personal names	Names of persons	Pelosi, Nancy, 1940–
Corporate names	Names of organizations or groups	Magnetic Fields (Musical group)
Titles	Names associated with works	Beowulf; Qur'an; Hamlet (Shakespeare, William)
Genre/Form	Types or kinds of bibliographic structures, literary or artistic forms, and formats	Science fiction television programs; Biography

Although faceted, FAST *does* allow for some subdivisions, but each facet's main headings can be subdivided only by subdivisions from the same facet. Thus, topical headings may be subdivided only by topical subdivisions (e.g., **Hospitals—Administration**) but not by a chronological or any other type of subdivision. Each of the headings and heading-with-subdivision combinations is established with its own authority record in the FAST authority file.

ONTOLOGIES

Ontologies are not *just* another type of controlled vocabulary. Although they are similar to thesauri in that they address relationships among concepts, there are significant differences. The most notable is that the purpose of an ontology is farther reaching than that of a controlled vocabulary. A thesaurus, most often, is focused on addressing synonyms, homographs, hierarchy, and so on. Although an ontology is a formalized vocabulary of terms, an ontology's emphasis is not solely lexical; it is also an attempt to represent the reality or essence of a situation, knowledge domain, or conceptual framework through formal definitions and specific relationships (i.e., those going beyond simple hierarchy and synonymy). In its attempt to provide a detailed understanding of a particular knowledge domain or subject area, an ontology is like a cross between a controlled vocabulary and a conceptual model. LIS professor Elin Jacob states,

> Ontologies have been variously construed as classification schemes, taxonomies, hierarchies, thesauri, controlled vocabularies, terminologies and even dictionaries. While they may display characteristics reminiscent of each of these systems, to equate ontologies with any one type of representational structure is to diminish both their function and their potential in the evolution of the Semantic Web.[44]

Ontology is a word with many nuances, which is ironic considering the common purpose of the term. Definitions of *ontology* are not always in alignment nor are they always clear; some of the more precise ones are presented here. In the field of philosophy the term *ontology* has a long and respectable history meaning a systematic account of existence. More recently the term has come to be used in the information science community for the categories of things that may exist in a particular domain and to refer to the knowledge shared by persons working in that domain; in other words, it is a systematic account of the entities and their relationships found in a particular shared reality to enable computers to make inferences or deductions based on the clearly defined relationships among entities. According to computer scientist Thomas Gruber, "An ontology defines (specifies) the concepts, relationships, and other distinctions that are relevant for modeling a domain. The specification takes the form of the definitions of representational vocabulary (classes, relations, and so forth), which provide meanings for the vocabulary and formal constraints on its coherent use."[45] According to Marcia Lei Zeng and Jian Qin, an ontology describes

(a) the types of things that exist (*classes*),
(b) the relationships between them (*properties*), and
(c) the logical ways those classes and properties can be used together (*axioms*).[46]

Jacob states, "Following Gruber's lead, an ontology can be defined as a partial, simplified conceptualization of the world as it is assumed to exist by a community of users—a conceptualization created for an explicit purpose and defined in a formal, machine-processable language."[47]

The broadest usage of the term is for a formal representation or specification of what is common sense or objective reality to a human being. An ontology defines the nature of reality by identifying the classes, properties, individuals, and data values in a particular knowledge domain in order to model the relationships among them, with the ultimate goal of making web content more accessible to machines.[48] It is created to keep conceptual and semantic ambiguity at a minimum in an information and technological environment, which is something that is not always possible with traditional controlled vocabularies. For example, in the *NASA Thesaurus*,[49] a traditional controlled vocabulary, you find the following concepts to help describe resources about the first moon landing. The terms are placed in hierarchies, which represent broader and narrower term relationships:

celestial bodies
  natural satellites

moon

personnel
flying personnel
astronauts

space flight
manned space flight
Apollo flights
Apollo 11 flight

You also find a small number of related terms.

Apollo 11 flight
RT Earth-Moon trajectory
lunar flight
lunar landing
manned spacecraft
astronauts
RT astronautics
crews
pilots
spacecrews

Although the relationships among them are not explicitly defined, there is no doubt that these terms and the vocabulary's syndetic structure can be helpful when retrieving resources within a database or catalog. The use of standardized terms and relationship indicators allows for successful searching and navigation. On the other hand, to better reflect reality in order to help machines "understand" and process information about this knowledge domain, additional entities and relationships are necessary. Entities such as *Buzz Aldrin*, *Neil Armstrong*, *Michael Collins*, *command module*, *lunar module*, and so on, are needed to bring out a more robust, realistic, and accurate picture of the event (i.e., domain) being modeled. Not only that, but additional relationships, such as *commanderOf/commandedBy*, *componentOf/containsComponent*, *landedOn/landingSiteOf*, *pilotOf/pilotedBy*, and so forth, are necessary to achieve the specificity needed for machine processing and inferences.[50] This is an example of how ontologies are different from traditional controlled vocabularies in terms of scope and function.

A formal ontology is useful for enhancing interoperability among systems in different knowledge domains or for creating intelligent agents that can perform certain tasks; both of which are goals of the Semantic Web. The vision for the Semantic Web requires that terms have explicit meanings and relationships so that machines can automatically process information found online. Ontologies providing such meaning are building blocks for the Semantic Web. If online agents are to be able to search and/or merge information from diverse communities, then ontologies are important to that outcome. This is because the same term may be used in different contexts with different meanings, and the same meaning may be represented by different terms in different contexts. In order to standardize the means for creating ontologies that can be used online, the Web Ontology Working Group of the W3C developed the OWL Web Ontology Language.[51]

The relationships identified in OWL (and its second edition, OWL2) are somewhat different from those found in thesauri. According to the *OWL2 Primer*,[52] like any modeling language, OWL is concerned with *individuals* (i.e., entities, such as a person or a thing), *classes* (i.e., groups or categories of entities with something in common), and *properties* (i.e., characteristics or relationships). In addition, there are other concerns brought out in OWL. These include *class hierarchies* (e.g., one class is subclass of another), which allow inferences to be made based on the relationships among groups. If *Lassie* (individual) is a member of the class *dogs*, and *dogs* is a subclass of *animals* (another class), then a machine can deduce that Lassie is an animal, and what is true of all animals is true for Lassie, but what is true for Lassie is not necessarily true for all animals, based on class hierarchies. Individuals can be members of multiple classes (e.g., Lassie is a dog, but Lassie is also a *pet*). Some classes may be equivalent (e.g., *animals* and *beasts*), but some are not (e.g., *cats* and *dogs*). When membership in one class (e.g., *dogs*) excludes the entity from memberships in another class (e.g., *cats*), those classes are said to be incompatible (i.e., *disjoint*) classes. Beyond the class structures enumerated, OWL also describes individuals and their relationships through *properties* (e.g., Lassie *hasSister* Sassie). These statements can be made both in positive and negative forms. Property statements also can be hierarchically structured. Some relationships provided for in OWL are:

- **subClassOf—**x is a subset of the class y (e.g., the class *dogs* is a subset of the class *animals*)
- **oneOf—**the enumerated thing is an instance of the class (e.g., Lassie is one member of the class *dogs*)
- **equivalentClass—**two class descriptions have the same set of individuals (e.g., the class *animals* is equivalent to the class *beasts*)
- **intersectionOf—**a class can belong to two other classes (e.g., the class *dogs* can be both *animals* and *pets*)
- **disjointWith—**two classes have no individuals in common (e.g., the class *cats* does not include any individuals of the class *dogs*)

The relatively simple structures presented here are just the beginning and are intended only to illustrate some basic differences between ontologies and the other forms of controlled vocabularies discussed earlier. Ontologies can be quite complex, the details of which go beyond the scope of this book.

NATURAL LANGUAGE APPROACHES TO SUBJECTS

In addition to controlled approaches to subject terminology, there are also natural language or "uncontrolled" approaches. These uncontrolled approaches include the longstanding practices of Natural Language Processing (NLP) and keyword searching, as well as the development of *tagging* (also known as *user tagging* or *social tagging*), which has been incorporated into a variety of web-based environments and a small number of information retrieval tools.

Natural Language Processing

Natural language processing (NLP) research aims to enable computers to interpret and react to human languages. So far, most progress in this area has focused on spoken and written language, though research into sign language is also being conducted.

Goals of NLP include machine translation, question answering, and creating conversational agents and dialog systems.[53] A goal of NLP, related to question answering, is improved and enhanced information retrieval (IR). The IR process can be described in three distinct steps: (1) to interpret users' information needs as expressed in free text, (2) to represent the complete range of meaning conveyed in documents, and (3) to "understand" when there is a match between the user's information need and all (and no more than) the documents that meet it. In order to accomplish any NLP task, certain challenges have to be addressed:[54]

- Sentences are often incomplete descriptions of what they mean. For example, "The door opened" does not tell whether the door was opened by a person, the wind, or its own weight. If the next sentence is, "Susan walked in," then the implication is that Susan opened the door, although that may or may not be the case.
- The same expression can mean different things in different contexts. For example, "Where's the water?" can mean that one is thirsty, or it can mean that one wants to know how to get to the beach.
- Natural languages are constantly gaining new words, usages, expressions, and meanings. During the 1998 Olympics one could hear that "The United States has not yet medaled." Most people had not heard the word *medal* used as a verb before that. Another example is the creation and evolution of the word *blog,* which descended from *web log* or *weblog* and can be used as both a noun and a verb.
- There are many ways to say the same thing. For example, "Mary registered for two summer courses" and "Mary signed up for two courses in the summer term" mean essentially the same thing. This problem applies not only to natural language sentences, but also to individual words, as discussed above with the challenges related to synonymous concepts.
- Sentences that are constructed identically can mean different things. In the two sentences, "Jennifer took the course with Professor Jones" and "Jennifer took the course with Mary," the first

sentence indicates that the professor taught the course that Jennifer took. But the second could mean that Jennifer and Mary are both students and took a course together, or it could mean that Professor Mary Jones likes to be addressed by her first name. Such ambiguities can often be sorted out through the context of surrounding sentences, but some cannot.

Successful NLP systems must contain information about the different areas of linguistics, as each area provides important information about how words, sentences, and dialogues are created and interpreted:[55]

- **Phonetics**: the study of how individual sounds are created and perceived. Not all languages have the same sounds; speech recognition systems need to be able to recognize different sounds depending on the language they are built for.
- **Phonology**: the study of patterns of sounds in languages, and how they combine to form syllables. Different languages have different "rules" for what phonemes can combine in what order and how a phoneme might change in different contexts. For example, in American English an *S* at the end of a word might be pronounced as a *Z*.
- **Morphology**: the study of the units of meaning (i.e., *morphemes*) in a language, and how they combine. Morphemes are words, prefixes, suffixes, and other word structures that can modify meaning. For example, in English, adding an *S* at the end of a word often indicates plurality.
- **Syntax**: the study of how words combine to create phrases and sentences and what combinations are allowed in different languages. Syntactic analysis includes identifying parts of speech and creating representations of the underlying structure of sentences.
- **Semantics**: the study of meaning in language, and the relationship between words and what they represent. For example, "She wants to print a web page" indicates a wanting event in which *she* wants a printing event to occur wherein she must have access to a web browser and a printer, but in order to understand that sentence an NLP system must have access to a knowledge base (i.e., an ontology) that includes what printing and a web page are.
- **Discourse analysis**: the study of how information is exchanged across sentences and how that context informs the semantic interpretation. For example, the meaning of pronouns such as *it, them,* and *her* can be given individual meanings only if what or who they refer to can be determined.
- **Pragmatics**: the study of how context affects interpretation. For example, the question "Do you have the time?" should not be answered with "*Yes,*" but should be interpreted as a request to be told the time. In the case of an IR system, the result of such analysis should be a translation to a command to be executed by the system. If the system is asked a question such as "Do you have anything on artificial intelligence?" the response should be a list of sources on artificial intelligence, not "*Yes.*"

Speech recognition systems must include knowledge of phonetics and phonology for the languages they are built for, in addition to the other linguistic fields. Text-based systems do not need to be concerned with the sounds of a language, but they have the additional tasks of identifying words and sentences, and interpreting punctuation, spelling, and other formatting or orthographic information that may add or change meaning in the text.

Keywords

One of the first approaches used by NLP researchers was the manipulation of keywords. The success of keyword searching depends upon at least two assumptions: (1) that authors writing about the same concepts will use the same words in their writings, and (2) that searchers will be able to guess what words those authors used for the concept. A 1993 study of journal articles in the pure sciences and social sciences looked at articles that shared common references.[56] An assumption was made that if articles shared common references, they dealt with the same or a related subject. It was found that few articles with common references shared common keywords. Another study the same year reported difficulty in choosing keywords for a literature search due to the use of multiple terms representing the same concepts.[57]

Among other problems discovered with keyword searching were that not all related information was retrieved, and searches often led to the extraction of irrelevant materials. A synonym list (i.e., synonym ring) approach was tried. Synonym lists were databases consisting of groupings of synonymous terms. When a keyword search was done, the synonym list was tapped to provide synonyms without the searcher having to be the one to think of all of them. This approach also failed for several reasons:

- the lists were not large or general enough;
- they were implemented in very small and specialized domains;
- the lists did not attempt any level of word role assignment (e.g., although one could substitute *aircraft* for *planes* and *big* for *large,* it was not possible to substitute *big aircraft* for *large planes* because the system had no knowledge of adjectives and nouns and which kinds of words could be used together to make a phrase).[58]

In IR systems that process full-text, the texts are analyzed and indexed when they are entered into the system and are retrieved through the use of keywords. Although full-text analysis retrieval systems are being marketed, many systems (e.g., distributed networks, the web) cannot transmit the complete text of hundreds of documents in response to a search query and do not adequately address the semantic problems described above. Sujata Banerjee and Vibhu Mittal proposed an indexing system using keyword searching combined with a well-developed lexical database, such as WordNet, to address some of these issues.[59] Banerjee and Mittal have proposed the following model (enhanced by this book's authors). For each keyword query (e.g., *family crisis*) of an IR system, results of an exact match would be given first. If there were insufficient matches the system would prompt the user for other options. It could first substitute synonymous adjectives from WordNet and then present a list to the user who would choose the combinations that make sense.[60] (The user would be involved also at each of the following stages.) Our example would yield the search terms: *household crisis, house crisis, home crisis,* and so forth. Then the system would substitute noun synonyms, resulting in a search for *family emergency.* The system would then drop adjectives, resulting in *crisis* being searched. Then the noun would be generalized to its hypernyms, resulting in: *family situation, family state of affairs, family juncture, family occasion.* Then the synonymous adjectives could be combined with the noun hypernyms, resulting in: *household situation, household state of affairs, home situation, home state of affairs,* and so forth. Also, the hyponyms of both noun and adjective could be searched, resulting in: *family challenge, family complication, family nightmare, foster home crisis, couple crisis, marriage crisis,* and so forth. Thus, there is potential for a very large lexical database, covering many fields of knowledge, to be used to enhance keyword searching of full-text documents (and possibly surrogate records) in IR systems that implement such tools. This, however, is not yet a common feature of most systems.

Tagging and Folksonomies

A recent venture into natural language approaches to the content of resources is *tagging.* This activity may occur within a contained system such as a catalog or on the open web. Tagging is a populist approach to description; it is a process by which a distributed mass of users applies keywords—also referred to as *tags* or *hashtags* (when formatted with the pound or hash [#] sign)—to various types of resources for the purposes of collaborative information organization and retrieval. Michalis Gerolimos states that the popularity of tagging "grew with the advent of social media and networking websites and brought an innovative element to what can generally be referred to as document description: users describing their own or someone else's documents or resources for personal (most of the time) purposes."[61] The use of tags is common on many web-based applications and websites that support interactive technologies. Although it is a popular approach to describing materials on the web, tagging is not a common feature in most retrieval tools found in LIS institutions. In 2012 Sharon Yang found that only 5 percent of integrated library systems and 47 percent of discovery interfaces allowed users to add tags to resources.[62]

Tagging allows individual users to group similar resources together by using their own terms or labels, with few or no restrictions. The tags assigned to a resource can be based on a variety of facets:

- **subject matter** (e.g., cooking, Sigur_Rós, metadata, PattyGriffin);
- **form** (e.g., images, humor, gossip, recipes);
- **purpose** (e.g., reference, delivery, travel, howTo);

- **time** (e.g., February, now, 2018, future);
- **location** (e.g., box3, bedroom, Albany);
- **tasks or status** (e.g., toRead, toDo, toSort, mine, own);
- **affective or critical reactions** (e.g., cool, fun, schlocky author, Questionable Literary Merit, yaas); or
- **myriad other reasons**, some of which may be unclear to others (e.g., I_am_morbid, woo-woo, G, 5, aargh).

The tags applied are then either used for searching, or they may be displayed in the form of an alphabetical list or as a tag cloud for browsing. Tag clouds are visual representations of all or only some of the tags assigned. In some systems, tag clouds display the tags used in the entire site, and in others, the tag clouds represent the tags of only one person or group. The tag cloud may look like a paragraph composed of individual words displayed in various font sizes—the font size representing the relative popularity of the tag. The larger the font is for a tag, the more that tag has been used in the system (see Figure 10.9). The individual tags in the cloud may be arranged alphabetically or ranked by the popularity of the tags, and in some systems, tags can be clustered into several mini-clouds. In other systems the tags may be displayed as a concept map laid out like a solar system, with one key concept found in the center and any number of satellites surrounding it.

Tagging may be applied in numerous domains to various types of resources. Tags may be used to label:

- web bookmarks (e.g., Delicious);[63]
- digital images and videos (e.g., Instagram,[64] Flickr,[65] and YouTube[66]);
- citations for scholarly articles (e.g., CiteULike[67]);
- products for sale (e.g., Amazon.com);[68]
- social network and blog postings (e.g., Twitter,[69] Facebook,[70] WordPress,[71] and Blogger[72]);
- materials in one's personal library (e.g., Library-Thing[73]); and
- individual resources in online catalogs and other retrieval tools (e.g., Ann Arbor District Library catalog[74] and University of Illinois at Urbana-Champaign Library catalog[75]).

According to Thomas Vander Wal, "The value in this external tagging is derived from people using their own vocabulary and adding explicit meaning. . . . People are not so much categorizing, as providing a means to connect items . . . to provide their meaning in their own understanding."[76]

The aggregation of these tags created by a large number of individual users results in what is referred to as a *folksonomy* (a blend of *folks* and *taxonomy*). Thomas Gruber states that with folksonomies, "we now have an entirely new source of data for finding and organizing information: user participation. . . . Tags introduce distributed human intelligence into the system."[77] Alex Wright states, "A folksonomy is a user-generated classification, emerging through bottom-up consensus a la Flickr, Furl, and Del.icio.us."[78] Emanuele Quintarelli states folksonomy is an activity of "collaborative categorization using freely chosen keywords by a group of people cooperating spontaneously."[79] She goes on to identify both drawbacks and strengths of folksonomy. She acknowledges that folksonomies lack precision in language, they lack syndetic structures, they do not work well in retrieval, and they do not scale well. On the other hand, she finds that folksonomies do reflect the general populace's language and needs, they are inclusive of everyone's input, they are helpful for understanding resources, and they are low cost.[80]

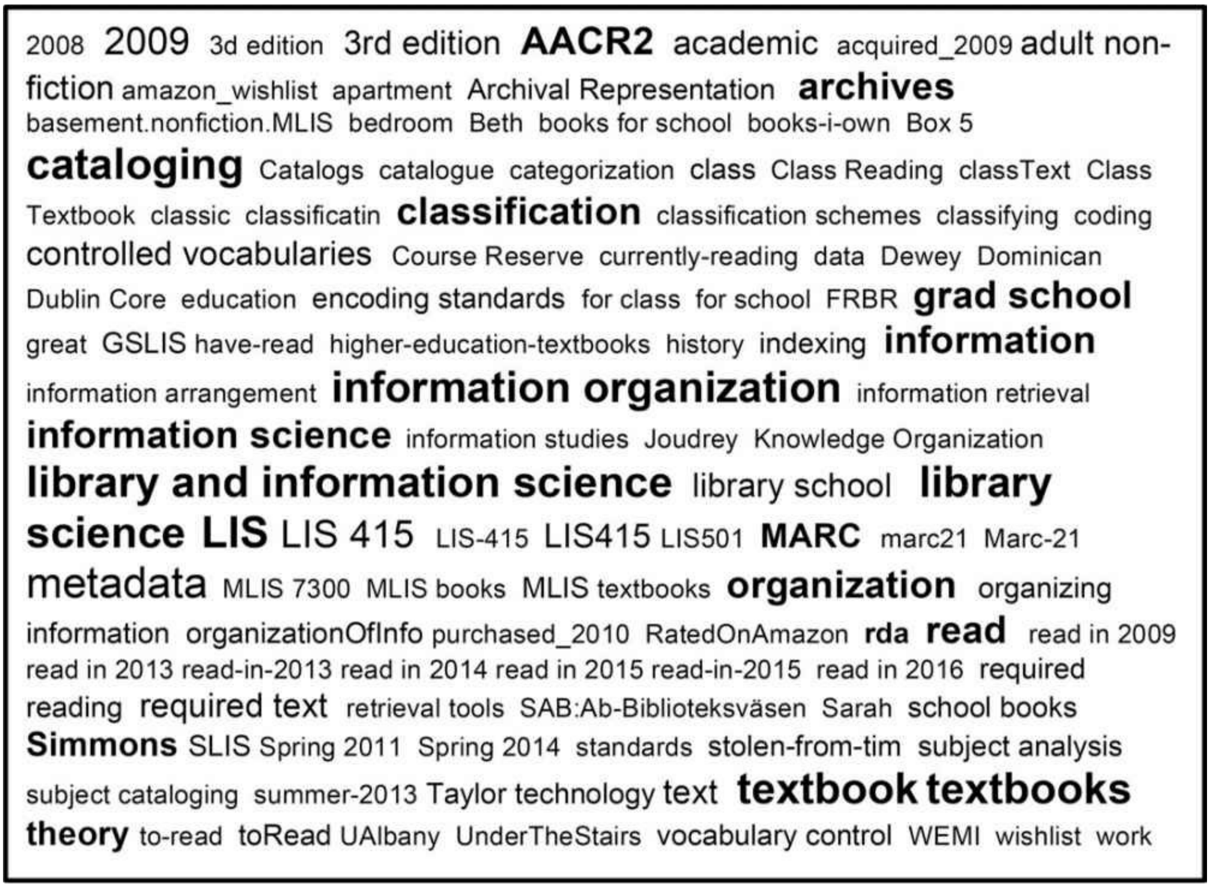


Figure 10.9. A Representation of a Tag Cloud for *The Organization of Information*, 3rd ed.

The implication is that if enough users tag enough resources, sufficient data can be aggregated to achieve stability, reliability, and consensus. In order for this data to be useful in augmenting current approaches to subject access, though, a critical mass of tags must be accumulated. The idea is that many tags applied to discrete resources by myriad individuals (who are tagging for countless reasons) will provide sufficient information to understand the nature of the resource and to allow us to take technological advantage of an inexpensive way to organize web-accessible information resources. Without a sufficient volume of data, however, the benefits are limited. For example, one or two user tags assigned to a catalog record may not provide extra insight into the resource being described or reliable supplementary access points; but 200 tags, similar in purpose, may provide additional, useful, meaningful machine-derived subject access points for the record.

Tagging is an exciting development to some because it is an approach to subject metadata by the people, for the people—without restrictions, unfamiliar jargon, and complex application rules. Users can assign as many tags as they like, and the terminology they use is their own. It can be done by non-experts, and it is primarily performed by volunteers. Although tagging may eventually provide added value to our current retrieval tools, there are also some drawbacks associated with user tagging. These problems are familiar to anyone who has ever performed a keyword search online or in another retrieval tool: no synonym and homograph control, no control of word forms (i.e., singular versus plurals, etc.), impaired precision and recall, and no hierarchical and associative relationships identified. For example, the genre *science fiction* can be represented by any of the small number of possible tags presented here:

- *science fiction*
- *sciencefiction*
- *science_fiction*
- *ScienceFiction*
- *scifi*
- *SciFi*

- *sci-fi*
- *Sci/Fi*
- *syfy*
- *sy-fy*
- *Sy Fy*
- *fiction.sciencefiction*
- *ciencia ficción*
- *sf*
- *SF*
- *sff*
- *scifi/fantasy*
- *siensfixion*

In other words, tagging lacks all of the benefits of controlled vocabularies. In addition, the tags assigned by some users may be so idiosyncratic or personal that they are of no real value to anyone else or may be misleading. Despite these shortcomings, tagging appears to be here to stay, at least on social media sites; the general populace has embraced it and besides being helpful, many find it fun to do. After more research into how tagging can be leveraged has been conducted, information professionals will need to find inexpensive ways of taking advantage of all of this user-generated metadata to supplement, augment, and enhance the expert-created subject metadata in the form of controlled vocabulary terms and classification notations.

CONCLUSION

This chapter has addressed issues of controlled vocabulary. Controlled vocabularies are often thought of as subject access tools, but they can be used for other purposes as well. All controlled vocabularies have to deal with issues and problems during their construction; understanding these issues contributes to making the best use of the vocabulary, as does understanding the general principles of application. Controlled vocabularies may come in several different forms: *simple term lists*, *synonym rings*, *taxonomies*, *thesauri* (including *subject heading lists*), and *ontologies* (although they are somewhat different). Ontologies have evolved as a way to help humans and computers share knowledge about a particular knowledge domain. Ontologies have grown out of the computer science community and from Natural Language Processing (NLP), which holds promise for sophisticated keyword approaches.

For subject access, subject heading lists and thesauri are the most prevalent types of vocabularies in use in information institutions today. Subject heading lists created by libraries were the first to appear. Thesauri are more strictly hierarchical and, for the most part, are developed in subject-specific situations and/or commercial indexing services. The latest addition to verbal approaches to subject access is the folksonomy, which is made up of user tags assigned to web-based resources. The next chapter is also concerned with subject access, but from the point of view of categorization and classification.

Some Important Terms in This Chapter

Abbreviation	Genre/Form Heading	Pre-coordination
Aboutness	Genus–Species Relationship	Preferred Term
Acronym	Hierarchical Relationship	Property
Antonym	Holonym	Qualifier Reference
Associative Relationship	Homograph	Related Term
Authorized Term	Homophone	See Also Note
Axiom	Hypernym	Simple Term List
Broader Term	Hyponym	Specific Entry
Class	Individual	Specificity
Class Hierarchy	Initialism	Subject Heading List
Classification	Instance Relationship	Subdivision
Coextensive Entry	Instruction Sheet	Syndetic Structure
Controlled Vocabulary	Keyword	Synonym
Coordinate Term	Literary Warrant	Synonym Ring
Cross-Reference	Meronym	Tag
Data Value Standard	Narrower Term	Tagging
Descriptor	Natural Language Processing	Taxonomist
Direct Entry	Ontology	Taxonomy
Entry Vocabulary	Pick List	Thesaurus
Equivalence Relationship	Post-coordination	Tree Structure
Folksonomy		Whole–Part Relationship

Some Important Acronyms in This Chapter

AAT: *Art & Architecture Thesaurus*
BT: Broader term
CONA: *Cultural Objects Name Authority*
ERIC: Education Resources Information Center
ET: *Ethnographic Thesaurus*
LC: Library of Congress
LCDGT: *Library of Congress Demographic Group Terms*
LCGFT: *Library of Congress Genre/Form Terms for Library and Archival Materials*
LCNAF: Library of Congress/NACO Authority File
LCSH: *Library of Congress Subject Headings*
MeSH: *Medical Subject Headings*
NLM: National Library of Medicine
NLP: Natural language processing
NT: Narrower term
OWL: Web Ontology Language
PSD: Policy and Standards Division (LC)
RDF: Resource Description Framework
RT: Related term
SA: See also
SAC: Subject Analysis Committee (ALA/ALCTS)
SHM: *Subject Headings Manual*
TGN: *Getty Thesaurus of Geographic Names*
UF: Used for
ULAN: *Union List of Artist Names*

NOTES

All URLs accessed June 2017.

1. David Crystal, *Internet Linguistics: A Student Guide* (London: Routledge, 2011), 98–103.

2. Thomas Mann, "'Cataloging Must Change!' and Indexer Consistency Studies: Misreading the Evidence at Our Peril," *Cataloging & Classification Quarterly* 23, nos. 3/4 (1997): 3–45.

3. National Information Standards Organization, *ANSI/NISO Z39.19-2005 (R2010): Guidelines for the Construction, Format, and Management of Monolingual Thesauri* (Baltimore, MD: NISO Press, 2010), 16–19, http://www.niso.org/apps/group_public/download.php/12591/z39-19-2005r2010.pdf.

4. "RDA Media Type: Concepts," Open Metadata Registry, http://metadataregistry.org/concept/list/vocabulary_id/37.html.

5. UNESCO, "UNESCO Thesaurus," http://vocabularies.unesco.org/browser/thesaurus/en/.

6. The American Folklore Society, "The AFS Ethnographic Thesaurus," http://www.afsnet.org/?page=AFSET.

7. National Aeronautics and Space Administration (NASA) Scientific and Technical Information (STI) Program, *The NASA Thesaurus*, https://www.sti.nasa.gov/nasa-thesaurus/.

8. ERIC Educational Resources Information Center, "Thesaurus," https://eric.ed.gov/.

9. ANSI/NISO Z39.19-2005 (R2010), 43–57.

10. Hans H. Wellish, "Aboutness and Selection of Topics," *Key Words* 4, no. 2 (March/April 1996): 9.

11. Sanford Berman, *Joy of Cataloging* (Phoenix, AZ: Oryx, 1981); *Prejudices and Antipathies: A Tract on the LC Subject Heads Concerning People* (Metuchen, NJ: Scarecrow, 1971), http://www.sanfordberman.org/prejant.htm.

12. Hope A. Olson, *The Power to Name: Locating the Limits of Subject Representation in Libraries* (Dordrecht, The Netherlands: Kluwer Academic, 2002).

13. For example, Joan K. Marshall, *On Equal Terms: A Thesaurus for Nonsexist Indexing and Cataloging* (Santa Barbara, CA: ABC-CLIO, 1977); Doris H. Clack, *Black Literature Resources: Analysis and Organization* (New York: Marcel Dekker, 1975).

14. Berman, *Prejudices and Antipathies*.

15. Steven A. Knowlton, "Three Decades since Prejudices and Antipathies: A Study of Changes in the Library of Congress Subject Headings," *Cataloging & Classification Quarterly* 40, no. 2 (2005): 127–28.

16. For an overview, see Lisa Peet, "Library of Congress Drops Illegal Alien Subject Heading, Provokes Backlash Legislation," *Library Journal*, June 13, 2016, http://lj.libraryjournal.com/2016/06/legislation/library-of-congress-drops-illegal-alien-subject-heading-provokes-backlash-legislation/.

17. Mann, "'Cataloging Must Change!'" 9–10.

18. Working Group on the Future of Bibliographic Control, "On the Record: Report of the Library of Congress Working Group on the Future of Bibliographic Control," Library of Congress, January 2008, http://www.loc.gov/bibliographic-future/news/lcwg-ontherecord-jan08-final.pdf.

19. Library of Congress, "Library of Congress Authorities," http://authorities.loc.gov/.

20. All three authority lists can be found at: The Getty, "Getty Vocabularies," http://www.getty.edu/research/tools/vocabularies/.

21. Princeton University, "WordNet: A Lexical Database for English," http://wordnet.princeton.edu/.

22. James D. Anderson and Melissa A. Hofmann, "A Fully Faceted Syntax for Library of Congress Subject Headings," *Cataloging & Classification Quarterly* 43, no. 1 (2006): 8–9.

23. Library of Congress Cataloging Policy and Support Office, "Library of Congress Subject Headings Pre- vs. Post-Coordination and Related Issues," http://www.loc.gov/catdir/cpso/pre_vs_post.pdf.

24. *Classification Web* (Washington, DC: Library of Congress, Cataloging Distribution Service), http://classificationweb.net/ [requires ID and password].

25. OCLC, *Connexion*, http://connexion.oclc.org/ [requires ID and password]. For membership information: https://www.oclc.org/en/membership/become-a-member.html.

26. Library of Congress, "Library of Congress Authorities," http://authorities.loc.gov/.

27. "Library of Congress Subject Heading PDF Files," Library of Congress, http://www.loc.gov/aba/cataloging/subject/.

28. *Subject Headings Manual*, 1st ed. (Washington, DC: Cataloging Distribution Service, Library of Congress, 2008). Also available in Cataloger's Desktop.

29. "Subject Heading[s] Manual PDF Files," Library of Congress, http://www.loc.gov/aba/cataloging/subject/.

30. *Cataloger's Desktop* (Washington, DC: Library of Congress, Cataloging Distribution Service), http://desktop.loc.gov/ [requires ID and password].

31. *Sears List of Subject Headings,* 21st ed., ed. Barbara A. Bristow (Ipswich, MA: Grey House, 2014).

32. The online version of *Sears* is available through EBSCOhost. See https://www.ebscohost.com/ for more information on subscriptions to this online product.

33. *Sears: Lista de Encabezamientos de Materia: Nueva Traducción y Adaptación de la Lista Sears*, ed. Iván E. Calimano (New York: H. W. Wilson, 2008).

34. National Library of Medicine, "Fact Sheet: Medical Subject Headings (MeSH)," https://www.nlm.nih.gov/pubs/factsheets/mesh.html.

35. National Library of Medicine, "Medical Subject Headings Browser," https://meshb.nlm.nih.gov/.

36. Library of Congress, "Introduction," *Library of Congress Genre/Form Terms for Library and Archival Materials*, https://www.loc.gov/aba/publications/FreeLCGFT/gftintro.pdf.

37. The Getty, "Art & Architecture Thesaurus Online: About the AAT," http://www.getty.edu/research/tools/vocabularies/aat/about.html.

38. The Getty, "The Art & Architecture Thesaurus Online," http://www.getty.edu/research/tools/vocabularies/aat/.

39. The Getty, "Still Life Painters," Art & Architecture Thesaurus Online, http://www.getty.edu/vow/AATFullDisplay?find=still+life+painter&logic=AND&note=&english=N&prev_page=1&subjectid=300266118.

40. Educational Resources Information Center (ERIC), *Thesaurus of ERIC Descriptors,* 14th ed. (Phoenix, AZ: Oryx, 2001).

41. ERIC Educational Resources Information Center, "Thesaurus," https://eric.ed.gov/.

42. Rebecca J. Dean, "FAST: Development of Simplified Headings for Metadata," in *Authority Control in Organizing and Accessing Information: Definition and International Experience*, eds. Arlene G. Taylor and Barbara B. Tillett (New York: Haworth Information Press, 2004), 331–32.

43. "SearchFAST," OCLC Research, http://fast.oclc.org/searchfast/.

44. Elin K. Jacob, "Ontologies and the Semantic Web," *Bulletin of the American Society for Information Science and Technology* 29, no. 4 (April/May 2003): 19, http://www.asis.org/Bulletin/Apr-03/BulletinAprMay03.pdf.

45. Tom Gruber, "Ontology," in *Encyclopedia of Database Systems*, eds. Ling Liu and M. Tamer Özsu (New York, Springer-Verlag, 2009), http://tomgruber.org/writing/ontology-definition-2007.htm.

46. Marcia Lei Zeng and Jian Qin, *Metadata*, 2nd ed. (Chicago: ALA, 2016), 61.

47. Jacob, “Ontologies and the Semantic Web,” 19.

48. W3C, “OWL 2 Web Ontology Language Document Overview (Second Edition),” 2012, https://www.w3.org/TR/owl-overview/.

49. NASA STI, *The NASA Thesaurus*, https://www.sti.nasa.gov/nasa-thesaurus/.

50. Example based on one mentioned briefly in Andreas Blumauer, “From Taxonomies over Ontologies to Knowledge Graphs,” *The Semantic Puzzle*, blog posting July 15, 2014, https://blog.semantic-web.at/2014/07/15/from-taxonomies-over-ontologies-to-knowledge-graphs/.

51. W3C, “OWL Web Ontology Language Overview,” http://www.w3.org/TR/owl-features/.

52. W3C, “OWL 2 Web Ontology Language Primer (Second Edition),” 2012, https://www.w3.org/TR/owl2-primer/.

53. Daniel Jurafsky and James H. Martin, *Speech and Language Processing*, 2nd ed. (Upper Saddle River, NJ: Pearson, 2009), 1–2.

54. Elaine Rich and Kevin Knight, *Artificial Intelligence,* 2nd ed. (New York: McGraw-Hill, 1991), 377–79.

55. Ibid., 379–80; Jurafsky and Martin, *Speech and Language Processing*, 2–3; and James Pustejovsky and Amber Stubbs, *Natural Language Annotation for Machine Learning* (Sebastopol, CA: O’Reilly Media, 2013), 3–4.

56. S. Nazim Ali, “Subject Relationship Between Articles Determined by Co-occurrence of Keywords in Citing and Cited Titles,” *Journal of Information Science* 19, no. 3 (1993): 225–31.

57. Renee B. Bush, “A Bibliography of Monographic Works on Biomaterials and Biocompatibility,” *Journal of Applied Biomaterials* 4, no. 2 (1993): 195–209.

58. Sujata Banerjee and Vibhu O. Mittal, “On the Use of Linguistic Ontologies for Accessing and Indexing Distributed Digital Libraries,” in *Digital Libraries ’94: Proceedings of the First Annual Conference on the Theory and Practice of Digital Libraries, June 19–21, 1994, College Station, Texas*, ed. John L. Schnase (College Station: Texas A&M University, 1994), http://www.csdl.tamu.edu/DL94/paper/banerjee.html.

59. Ibid.

60. Ibid.

61. Michalis Gerolimos, “Tagging for Libraries: A Review of the Effectiveness of Tagging Systems for Library Catalogs,” *Journal of Library Metadata* 13, no. 1 (2013): 37.

62. Sharon Yang, “Tagging for Subject Access: A Glimpse Into Current Practices by Vendors, Libraries and Users,” *Computers in Libraries* 32, no. 9 (2012): 20–21.

63. Delicious Social Bookmarking, https://del.icio.us/.

64. Instagram, https://www.instagram.com/.

65. Flickr, http://www.flickr.com/.

66. YouTube, http://www.youtube.com/.

67. CiteULike, http://www.citeulike.org/.

68. Amazon.com, http://www.amazon.com/.

69. Twitter, https://twitter.com/.

70. Facebook, https://www.facebook.com/.

71. WordPress, https://wordpress.com/.

72. Blogger, https://www.blogger.com/.

73. LibraryThing, https://www.librarything.com/.

74. Ann Arbor District Library catalog, http://www.aadl.org/catalog/.

75. I-Share Catalog, University of Illinois at Urbana-Champaign Library, https://vufind.carli.illinois.edu/vf-uiu/.

76. Thomas Vander Wal, "Folksonomy Coinage and Definition," http://vanderwal.net/folksonomy.html.

77. Thomas Gruber, "Ontology of Folksonomy: A Mash-Up of Apples and Oranges," *International Journal on Semantic Web and Information Systems* 3, no. 1 (2007): 3–4, http://tomgruber.org/writing/ontology-of-folksonomy.htm.

78. Alex Wright, "Folksonomy," *Alex Wright* blog archives, http://alexwright.org/blog/archives/000900.html.

79. Emanuele Quintarelli, "Folksonomies: Power to the People," a paper presented at the ISKO Italy-UniMIB Meeting, 2005, http://www.iskoi.org/doc/folksonomies.htm.

80. Ibid.

SUGGESTED READINGS

Banerjee, Sujata, and Vibhu O. Mittal. "On the Use of Linguistic Ontologies for Accessing and Indexing Distributed Digital Libraries." In *Digital Libraries '94: Proceedings of the First Annual Conference on the Theory and Practice of Digital Libraries, June 19–21, 1994, College Station, Texas,* edited by John L. Schnase. College Station: Texas A&M, 1994. http://www.csdl.tamu.edu/DL94/paper/banerjee.html.

Berman, Sanford. *Joy of Cataloging.* Phoenix, AZ: Oryx, 1981.

Berman, Sanford. *Prejudices and Antipathies: A Tract on the LC Subject Heads Concerning People*. Metuchen, NJ: Scarecrow, 1971.

Chan, Lois Mai. *Library of Congress Subject Headings: Principles and Applications.* 4th ed. Westport, CT: Libraries Unlimited, 2006.

Clark, Alexander, Chris Fox, and Shalom Lappin, eds. *The Handbook of Computational Linguistics and Natural Language Processing*. Malden, MA: Wiley-Blackwell, 2010.

Cleveland, Donald B., and Ana D. Cleveland. *Introduction to Indexing and Abstracting.* 4th ed. Santa Barbara, CA: Libraries Unlimited, 2013. Chapter 11: "The Indexing Process" and Chapter 15: "Indexing Specific Areas."

"ERIC's Indexing and Retrieval: 2001 Update" and "Thesaurus Construction and Format." In *Thesaurus of ERIC Descriptors,* 14th ed., edited by James E. Houston, xiv–xxxi. Phoenix, AZ: Oryx, 2001.

The Getty. "Art & Architecture Thesaurus Online: About the AAT." http://www.getty.edu/research/tools/vocabularies/aat/about.html.

Jacob, Elin K. "Ontologies and the Semantic Web." *Bulletin of the American Society for Information Science and Technology* 29, no. 4 (April/May 2003): 19–22. http://www.asis.org/Bulletin/Apr-03/BulletinAprMay03.pdf.

Joudrey, Daniel N., Arlene G. Taylor, and David P. Miller. *Introduction to Cataloging and Classification*. 11th ed. Santa Barbara, CA: Libraries Unlimited, 2015. Chapter 12: "Verbal Subject Access," Chapter 13: "*Library of Congress Subject Headings* (LCSH)," Chapter 14: "*Sears List of Subject Headings (Sears),*" and Chapter 15: "Other Verbal Access Systems."

Kao, Anne, and Stephen R. Poteet, eds. *Natural Language Processing and Text Mining*. London: Springer, 2007. Chapter 1: "Overview."

Lancaster, F. W. *Indexing and Abstracting in Theory and Practice.* 3rd ed. Champaign: University of Illinois, Graduate School of Library and Information Science, 2003. Chapter 2: "Indexing Principles," Chapter 4:

"Precoordinate Indexes," Chapter 5: "Consistency of Indexing," Chapter 6: "Quality of Indexing," Chapter 16: "Indexing and the Internet," and Chapter 17: "The Future of Indexing and Abstracting."

Library of Congress, Catalogers Learning Workshop. "Library of Congress Subject Headings: Online Training." Janis L. Young and Daniel N. Joudrey, instructors. https://www.loc.gov/catworkshop/lcsh/.

Mann, Thomas. *The Oxford Guide to Library Research.* 4th ed. New York: Oxford University Press, 2015. Chapter 2: "Subject Headings and the Library Catalog," Chapter 4: "Subject Headings and Descriptors in Databases for Journal Articles," and Chapter 5: "Keyword Searches."

National Library of Medicine. "Welcome to Medical Subject Headings!" https://www.nlm.nih.gov/mesh/.

Noy, Natalya F., and Deborah L. McGuinness. *Ontology Development 101: A Guide to Creating Your First Ontology.* Knowledge Systems Laboratory, Stanford University, 2000. http://www.ksl.stanford.edu/people/dlm/papers/ontology-tutorial-noy-mcguinness-abstract.html.

Olson, Hope A. *The Power to Name: Locating the Limits of Subject Representation in Libraries*. Dordrecht, The Netherlands: Kluwer Academic, 2002.

Pustejovsky, James, and Amber Stubbs. *Natural Language Annotation for Machine Learning*. Sebastopol, CA: O'Reilly Media, 2013.

Quintarelli, Emanuele. "Folksonomies: Power to the People." Paper presented at the ISKO Italy-UniMIB Meeting, 2005. http://www.iskoi.org/doc/folksonomies.htm.

Svenonius, Elaine. *The Intellectual Foundation of Information Organization.* Cambridge, MA: MIT Press, 2000.

CHAPTER 11
SYSTEMS FOR CATEGORIZATION

Chapter 10 discusses methods of providing subject access using words or tags. Another type of subject access is provided through categories. Both approaches attempt to translate the aboutness of information resources into an artificial subject language and to bring together like materials, but each approach does this in a different way. When using verbal techniques, topical words or tags are assigned to a resource; in systems for categorization, a resource is placed into the most suitable group available. These groupings, which may or may not be represented by notations or symbols, are often based on disciplines or broad forms of knowledge.

Categorization has a much longer history than that of controlled vocabularies. Philosophers have tried to categorize knowledge for many centuries. Classification is a form of categorization, but over the last two centuries classification (particularly bibliographic classification) has come to be associated with assigning some kind of notation to physical information resources; this is reinforced by the tendency of many to confound the ideas of classification notations and call numbers. (Call numbers are not addressed in this chapter. For more information about call numbers, see Appendix B.) In the thinking of some, the connection to categorization has been lost. However, the categories devised by philosophers in past centuries are the bases for the major classification schemes still in use today.

Classification theory has not received as much attention in the United States as it has in other places around the world such as India, Great Britain, and some other European countries. There has been a tendency in the United States to see classification only as a location device for the arrangement of physical resources on a shelf (exemplified by the phrase "mark 'em and park 'em"). Nevertheless, there is great potential for categorizing all resources, including online ones. This chapter addresses this potential along with the following questions: What are categories, classifications, and taxonomies? How do humans categorize? How is subject content expressed through symbols or notation? What conflicts arise in the application of classification to resources?

WHAT ARE CATEGORIES, CLASSIFICATIONS, AND TAXONOMIES?

The concepts *classification* and *categorization* are often used interchangeably, but are they the same? The answer is not clear. This question is further complicated by the resurgence of the term *taxonomy* for systems of categorization that are used online and in intranet systems. Before moving on to various approaches to grouping similar resources or ideas, it may be helpful to attempt to discuss and distinguish among the terms *categorization, classification,* and *taxonomy.*

Here are some definitions provided by the *Oxford English Dictionary* (OED):

- *Categorization*: to place in a category or categories; to classify
- *Category*: a term . . . given to certain general classes of terms, things, or notions; a class, or division, in any general scheme of classification
- *Classification*: the action of classifying or arranging in classes, according to shared characteristics or perceived affinities; a systematic distribution, allocation, or arrangement of things in a number of distinct classes.
- *Class*: a set or category of things having some related properties or attributes in common, grouped

Merriam-Webster defines *classification* as the "systematic arrangement in groups or categories according to established criteria."[2] Ultimately, neither source helps to distinguish between classification and categorization.

This lack of clarity is not confined only to dictionaries. The two words *classification* and *categorization* are sometimes used interchangeably in the library and information science (LIS) literature. Some, however, are more rigorous in their use of the terms. Classification has specific connotations in libraries; thus, over the years some distinctions have been made. In general, categorization is considered to be broader and more abstract than classification; categorization is a cognitive function that is used to group concepts, rather than a structured process used to systematically arrange physical resources. Categorization can be seen as an amorphous or less well-defined grouping; whereas classification can be viewed as a comprehensive hierarchical structure for organizing information resources on linear shelves. Elin Jacob points out that in the LIS field the term *classification* can have several senses. The term may refer to three distinct but related concepts:

- a system of classes, ordered according to a predetermined set of principles and used to organize a set of entities;
- a group or class in a classification system; and
- the process of assigning entities to classes in a classification system.[3]

Jacob also compares *categorization* and *classification*:

> Although systems of classification and categorization are both mechanisms for establishing order through the grouping of related phenomena, fundamental differences between them influence how that order is effected—differences that do make a difference in the information contexts established by each of these systems. While traditional classification is rigorous in that it mandates that an entity either is or is not a member of a particular class, the process of categorization is flexible and creative and draws nonbinding associations between entities—associations that are based not on a set of predetermined principles but on the simple recognition of similarities that exist across a set of entities.[4]

Commenting on a 1977 conference paper by P. A. Studer,[5] Christopher Dent also attempts to distinguish between the terms *classification* and *categorization* to countermand the lack of precision found in the LIS literature:

> In my view *classification* is an artificial (synthetic, non-fundamental) process by which we organize things for presentation or later access. It involves the arbitrary creation of a group of classes, which have explicit definitions and may be arranged in a hierarchy. In other words, a class is strictly defined and once inhabited the inhabitants can be enumerated. *Categorization* on the other hand is a natural process in the sense that humans do it as part of their cognitive fundament. It is, like Studer reports, an act of simplification to make apprehension and comprehension of the environment more efficient. Categories spring up out of necessity and because they are designed to replace the details of definition are themselves resistant to definition.[6]

To put it more succinctly, Dent concludes, a class "is a defined grouping of entities in which the members fulfill the definition of the class and can be listed" and a category "is a cognitive label applied to a non-enumerable grouping of entities wherein membership is determined by typicality amongst the members and not some overarching definition."[7] Although few are as precise as Dent or Jacob in differentiating between the two concepts, these distinctions are helpful in understanding the discussions of categories and classes that follow.

Complicating this discussion is the renewed popularity of the term *taxonomy*. Definitions of taxonomy abound in the literature, and some of those definitions contradict each other. Amy Warner says that the terms *taxonomies, thesauri,* and *classification systems* are synonyms. They are "organized lists of words and phrases, or notation systems, that are used to initially tag content, and then to find it through navigation or search."[8] She states that definitions of these terms are currently in flux and hopes that definitions will soon be standardized. Similar confusion with definitions is indicated in the *Montague Institute Review*: "A taxonomy is a system for naming and organizing things into groups that share similar characteristics."[9] These definitions do not include the necessity for hierarchy, but Thomas Wason says that, "A taxonomy is a knowledge map of a topic, typically realized as a controlled vocabulary of terms and or phrases. A taxonomy is an orderly classification of information according to presumed natural relationships. . . . The most typical form of a taxonomy is a hierarchy."[10] F. W. Lancaster, in decrying the "rediscovery of wheels" he saw in "new" concepts wrote:

> My biggest complaint, however, is the fact that the noun *classification* has virtually been replaced by (shudder!) *taxonomy,* (double shudder!!) *ontology,* or even (triple shudder!!!) *taxonomized set of terms.* The way these terms are defined in recent articles clearly shows that they are used synonymously with *classification scheme.*[11]

The word *taxonomy* comes from the Greek *taxis* (arrangement, order) and *nomos* (law). Taxonomies have existed in the strict hierarchical world of science since Aristotle, in *Historia Animalium,* created a taxonomy of the animal kingdom. The most famous scientific taxonomy, however, is Linnaeus's *Systema Naturae,* a classification of plants and animals from 1735. The taxonomies being constructed today for use in intranets, though, do not necessarily follow such strict rules, nor are they always hierarchical. What they do seem to have in common is categories. Therefore, the discussions of categorization in this chapter are also applicable to taxonomies.

Lists of taxonomies that one finds online often include traditional classification schemes, subject heading lists, folksonomies, and ontologies, as well as subject-specific tools that actually call themselves *taxonomies*. Examples of the last are the Taxonomy of Educational Technology[12] and GRIN Taxonomy[13] from the U.S. National Plant Germplasm System. A number of taxonomies are proprietary to the organizations that created them and are under copyright, so they may not be freely used in other situations.

THEORY OF CATEGORIZATION

It was mentioned in Chapter 1 that human beings seem to have a basic drive to organize and that children begin to categorize early in their development. For example, it is essential that children learn to distinguish between things that are and are not edible. At later stages, children begin to categorize toys and other objects by color, shape, purpose, or their own notions of value. Adults have a great need for categories too. Categories abound in daily life—from the grocery store to the office to our homes. Most people, although they may not even realize it, categorize the various items found in their homes and offices every day; for example, few would place their forks and knives in the living room, because eating utensils are associated with kitchen items—not with couches, bookcases, or photos of grandma.

Theories of categories go back at least as far as the ancient Greeks. For example, the ancient philosopher and mathematician Pythagoras categorized the cycle of life as consisting of *birth*, *growth*, *decay*, *death*, *absorption*, and *metamorphosis*. Later, Empedocles placed every physical thing into one of four elements: *earth*, *air*, *fire*, and *water*. The most influential discussion of categories comes from Aristotle and his attempt to organize the objects and ideas of the world.

The Rise and Fall of the Classical Theory of Categories

The roots of contemporary classification systems can be traced back to Aristotle's "classical theory of categories." The word *category* comes from the Greek *kategorein* meaning *to accuse*, *to assert*, *to predicate*.[14] Aristotle's theory reflected this definition; his categories were 10 states of being or 10 things that can be expressed about an object or an idea. His categories are

- Substance
- Quantity
- Quality
- Relation
- Place
- Time
- Position
- State

- Action
- Affection[15]

Aristotle's theory placed objects or ideas into the same category based on what they have in common. This approach went unchallenged until the mid-twentieth century because, as George Lakoff says, categories were thought to be well understood. Until then, a category was considered to be an abstract container with strict binary membership, that is, things belonged either inside or outside of the container; there were no grey or fuzzy areas. Categories were also defined as mutually exclusive groups, that is, things could belong to one and only one category, and the shared common attributes of the group members were what defined the category.[16] This narrow view of categories, however, began to change about 65 years ago.

Cracks in the Classical Theory of Categories

A brief history of the research of the second half of the twentieth century will serve to illuminate the process of categorizing and its relationship to classification. This history is summarized from Lakoff's *Women, Fire, and Dangerous Things: What Categories Reveal about the Mind.*[17] Cracks began to appear in the classical theory of categories in 1953, when Ludwig Wittgenstein showed that a category like *game* does not fit the classical mold. This category has no single collection of common properties. For example, a game may be for education, amusement, or competition, and it may involve luck or skill. The category also has no fixed boundary, because new kinds of games can be added to it. Wittgenstein proposed that since there is no one attribute that is common to all games, "family resemblances" (a complex network of similarities and relationships) unite *games* into what we call a category. Just as members of a family have similarities, so do games.

J. L. Austin, in a paper published in 1961, extended Wittgenstein's analysis to the study of words. He wondered why we call different things by the same name (e.g., foot of a mountain, foot of a list, person's foot). Should not *mountain, list,* and *person* be in the same category if they all have a foot? But there are times when words do belong in the same category even though they share no common properties (e.g., *ball, bat,* and *umpire* can all go into the category *baseball*). Austin, like Wittgenstein, helped to show that traditional views of categories were inadequate.

Lotfi Zadeh contributed *fuzzy set theory* to the chipping away of the classical theory of categories. He noted that some categories are well defined while others are not. One either is or is not a member of a club, but whether one is tall or not depends to some extent upon the observer. The category *tall* is graded—one may be neither clearly tall nor clearly short; and a shorter person, looking at someone of medium height, might say that the person is tall. In 1965 Zadeh devised a form of set theory to include gradations of membership in categories.

Floyd Lounsbury's studies of Native American kinship systems also chipped away at the classical theory. He found that among various groups the same name (category) is used to express the kinship relationships of several different types of relatives. For example, in one group, uncles, great-uncles, and nephews on one's mother's side of the family are all called by the same word. (Perhaps they find it strange that in the dominant culture there are no verbal distinctions between maternal uncles, paternal uncles, or uncles by marriage.) The challenge to classical theory is that what seem to be definite and distinct categories in one culture and language are not the same categories in another culture and language.

Brent Berlin and Paul Kay (also mentioned in Chapter 9) published their work on color in 1969. They found that there could be anywhere from 2 to 11 expressions for the basic colors in different languages. (In recent years, some have identified a 12th basic color in various languages, but the color is not always the same!)[18] Although most people around the world can physiologically perceive and conceptually differentiate all of the 11 to 12 basic colors, depending on their language, they may not be able to express all of them as separate words. For example, in traditional Welsh, the word *glas* was used for blue but also for some shades that English speakers might refer to as *green* or *grey*. Again, it is shown that language and culture play a major role in the establishment of categories. Paul Kay and Chad McDaniel in 1978 described some follow-up research on colors. They drew on work that had been done in neurophysiology to conclude that human biology influences perception of color. They drew on fuzzy set theory to determine that in all cultures that have fewer than 11 basic colors, *cold* colors always include green, blue, and black, whereas *warm* colors always include red, orange, yellow, and white. Thus, color categories are not quite as arbitrary as some might infer from the Berlin and Kay studies.

Roger Brown began the study of basic level categories. His work published in 1965 observed that there is a first level at which children learn categories (e.g., *flower,* for any variety of flower). Yet, there are many names that can be used for such categories; some are more specific (e.g., *rose, daffodil)* and some more general (e.g., *plant*). Brown considered the children's level to be the "natural" level, whereas the more specific and general levels were viewed as "achievements of the imagination." Brent Berlin and associates, in research on naming plants and animals published from 1969 to 1977, also showed that there seems to be a universal level at which humans name things, and for plants and animals it is more likely to be at the genus level (e.g., *oak,* not *tree* and not *Sawtooth oak*; although this might not hold true for someone with experience only in an urban environment or for someone whose training has led to a more precisely honed level). It can be suggested that certain basic levels of categories have to do with being human and are the same across cultures.

Prototype Theory

The major crack in the classical theory of categories came when Eleanor Rosch developed *prototype theory* with her work between 1973 and 1981. She theorized that if, as classical theory states, categories are defined only by properties that all members share, then no members should be better examples of the category than any other members. She further theorized that if categories are defined only by properties that all members share, then categories should be independent of the humans doing the categorizing. She found that, contrary to classical theory, categories *do* have best examples (i.e., *prototypes*). For example, Rosch found in her research that people thought that *robin* was a better example of *bird* than was *ostrich.* And she found that human capacities *do* play a role in categorization (e.g., for someone 5 feet tall, there are many more tall people in the world than there are for someone 6 feet tall). This is related to Zadeh's *fuzzy set theory* mentioned earlier. *Ad hoc* categories also figure in here. Ad hoc categories are those that are made up on the spur of the moment. Different people will put different things into a category such as *camping gear* depending upon their experience, where they are going, how they will camp, and so forth.

Because the most widely used classification schemes in the United States are based upon the classical theory of categories, classifiers using them are sometimes quite frustrated to find that they have a subject concept that does not fit neatly into one of the categories. For example, multifaceted, cross-disciplinary information resources cannot easily be accommodated by the structure and rigidity of classical categories. These classifiers are often relieved to learn that classical theory has major cracks; so it is not unusual to run into non-prototypes and fuzzy sets.

BIBLIOGRAPHIC CLASSIFICATION

Classification, as noted earlier, is a structured system of categories used to collocate similar ideas or objects. One particular type of classification is *bibliographic classification,* which came into being for the purpose of arranging and retrieving information resources and later became used for arranging metadata records in library catalogs and other information retrieval tools. Bibliographic classification schemes (henceforth referred to as *classification schemes* or *classifications*) in the form that we know them are relatively new in the history of information organization. Early depositories of recorded information usually had some arrangement: title, broad subject, chronology, author, order of acquisition, or size. At the Alexandrian library, Callimachus's *pinakes* had at least 10 broad categories (or main classes). Arrangement within the classes tended to be by author. This seemed to be a model for arrangement of catalogs and bibliographies into the Early Middle Ages.

During the High Middle Ages, when monastery libraries became the keepers of books, there was little need for classification because libraries were so small. In the Late Middle Ages, the universities that developed began to divide their books according to the Trivium (Grammar, Rhetoric, Logic) and the Quadrivium (Arithmetic, Music, Geometry, Astronomy), the seven subjects taught in the medieval university. But within the seven classes, the books had fixed shelf locations.

Starting in the sixteenth century, librarians devised many different classification schemes. Often these were based upon philosophers' systems of knowledge. However, none caught on, and fixed locations continued to predominate. With the rapid growth of libraries in the nineteenth century, librarians felt a need for better arrangement so that the content of the collections would be more apparent to users. Philosopher Francis Bacon in the early seventeenth century had divided knowledge into three basic "faculties": history (natural, civil, literary, ecclesiastical); philosophy (including theology); and works of imagination (poetry, fables, etc.). This scheme had widespread influence, and numerous classification schemes were based upon it.

The most famous was that of Thomas Jefferson, who classified his own library before he eventually sold it to the Library of Congress (LC) as the basis of a new collection in 1815.

Some history of the development of specific classification schemes in the nineteenth and twentieth centuries is discussed in Chapter 2 and need not be repeated in full here. Very briefly, in 1876 Melvil Dewey published what later became known as the *Dewey Decimal Classification* (DDC), and shortly thereafter, Charles Ammi Cutter began work on his *Expansive Classification* (EC). Paul Otlet and Henri La Fontaine began development of the *Universal Decimal Classification* (UDC) in 1895, based on the fifth edition of DDC. At the beginning of the twentieth century, LC created its own classification (LCC), based loosely on the main class outline of Cutter's *Expansive Classification.* S. R. Ranganathan created his *Colon Classification* (CC) in the early 1930s and adapted the word *facet* as a term to indicate the various subparts of the whole classification. From the early 1970s and continuing through today, the Bliss Bibliographic Classification Association has been developing the second edition of the *Bibliographic Classification* (BC2); it is a total reworking of the classification created by Henry Bliss in the first half of the twentieth century and published in four volumes in the United States between 1940 and 1953. BC2 is being so thoroughly revised that it is considered to be an entirely new scheme, based on the faceting principles developed over much of the twentieth century by Ranganathan and the Classification Research Group (CRG) in the United Kingdom.

While classification is mostly associated with arranging tangible resources on shelves, it is also a useful way to divide large databases. It offers an alternative to alphabetical subdivision. And if several databases choose the same classification scheme, it is possible to enable subject searching across systems. Classification can also be used as a switching language among different languages. The notations of a classification are not language specific, and therefore the meanings of the notations can be given in whatever language is appropriate for the setting. For the DDC, for example, translations, either recently completed or in progress, include those in Arabic, Chinese, French, German, Greek, Hebrew, Icelandic, Italian, Korean, Norwegian, Russian, Spanish, Swedish, and Vietnamese, with an additional 15 older translations also available. The UDC is also multilingual; there are recent printed full editions in English, French, Portuguese, and Spanish, with online partial editions in over 50 more languages. The LCC has not yet moved in a multilingual direction.

Hierarchical, Enumerative, and Faceted Classifications

As most of the schemes mentioned above were devised before anyone challenged the classical theory of categories, these schemes are firmly based in *hierarchical* arrangements (even the faceted CC and BC2 are hierarchical within their individual sub-categories). Hierarchical schemes begin with broad, top-level categories, which branch into any number of subordinate levels, moving from the general to the specific (creating the familiar tree structure associated with hierarchies). For example, DDC starts with 10 main classes, divides each of those into 10 (creating 100 divisions), divides each of the divisions into 10 (creating 1,000 sections), divides each of those into 10 subsections (10,000), and so on into potential infinity (see Figures 11.1 and 11.2).

The same classification schemes are also *enumerative.* Enumerative schemes attempt to assign a designation for every subject concept (both single and composite) needed in the system. All schemes have elements that are enumerative, but some have much more than others. LCC is much more enumerative than DDC because it attempts to list a notation for every possible topic (or to allow for alphabetical arrangement of unlisted topics) within its 21 main classes (see Figure 11.3). Although DDC lists some basic numbers in the main part of its classification (called the *schedules*), it allows for more complex numbers to be constructed through the use of tables; LCC, on the other hand, includes this level of detail directly in its schedules (see Figure 11.4). UDC, originally based on DDC, was hierarchical and enumerative at its base when it was designed; it has now developed into a scheme that is somewhat closer to what is called a *faceted* scheme through the addition of number-building techniques.

Faceted classification, like hierarchical/enumerative schemes, attempts to include all possible subjects, but it does not do this by creating a singular place in a hierarchy for each topic (with its own specified number). Faceted classification is made up of many discrete topics. It is an attempt to divide the universe of knowledge into its component parts and then to gather those parts into individual categories or *facets*. Within each facet, the topics are assigned an individual notation. When the cataloger assigns a classification notation to an information resource, the notations assigned to each subtopic in the resource are identified and then strung together, in the appropriate order, to create a multidimensional classification. If one thinks of each of the faces of a cut and polished diamond as a facet of the whole, one can picture a classification notation that has small notations standing for subparts of the whole topic, which are pieced together to create a complete classification notation. This concept of faceted classification was named first by Ranganathan in his explanation of his *Colon Classification*. CC provides lists of symbols for single concepts, with rules for combining them into complex concepts. This approach is also referred to as *analytico-synthetic classification*, because the classification is established through an analysis of topics into component parts and then notation is synthesized from those parts.

Dewey Decimal Classification

10 Main Classes

000	Computer science, information & general works
100	Philosophy & psychology
200	Religion
300	Social sciences
400	Language
500	Science
600	Technology
700	Arts & recreation
800	Literature
900	History & geography

An Example of the Hundred Divisions

600	Technology	
600		Technology
610		Medicine & health
620		Engineering
630		Agriculture
640		Home & family management
650		Management & public relations
660		Chemical engineering
670		Manufacturing
680		Manufacture for specific uses
690		Construction of buildings

An Example of the Thousand Sections

600	Technology		
640		Home & family management	
640			Home and family management
641			Food and drink
642			Meals and table service
643			Housing and household equipment
644			Household utilities
645			Household furnishings
646			Sewing, clothing, management of personal and family life
647			Management of public households (Institutional housekeeping)
648			Housekeeping
649			Child rearing; home care of people with disabilities and illnesses

Figure 11.1. An Illustration of Hierarchical Arrangement in the *Dewey Decimal Classification* (DDC). (Source: WebDewey, OCLC)

600	Technology
640	Home & family management
640	Home and family management
641	Food and drink
641.2	Beverages (Drinks)
641.3	Food
641.4	Food preservation and storage
641.5	Cooking
641.51	Beginner and gourmet cooking
641.52-641.54	Cooking specific meals
641.52	First meal of the day
641.53	Light meals
641.54	Main meal of the day
641.55	Money-saving and timesaving cooking
641.56	Cooking for special situations, reasons, ages
641.57	Quantity, institutional, travel, outdoor cooking
641.58	Cooking with specific fuels, appliances, utensils
641.59	Cooking characteristic of specific geographic environments, ethnic cooking
641.6	Cooking specific materials
641.7	Specific cooking processes and techniques
641.8	Cooking specific kinds of dishes and preparing beverages

Figure 11.2. An Illustration of Additional Subdivisions Found in the DDC at 641.5. (Source: WebDewey, OCLC)

Ranganathan posited five fundamental categories that can be used to illustrate the facets of a subject. They are:

- **Personality (P)**: the focal or most specific aspect of the subject;
- **Material** or **Matter (M)**: a component;
- **Energy (E)**: an activity, operation, or process;
- **Space (S)**: a specific or generic place or location; and
- **Time (T)**: a chronological period, a year, a season, etc.

(These categories are often referred to by the initialism, *PMEST*.) For example, "the design of wooden furniture in eighteenth-century America" has all five facets. Furniture is the focal subject, or personality; wood is the material facet; design is an activity and so constitutes the energy facet; the space is America; and the time is the eighteenth century. If each facet has a specific notation (e.g., let's say 28 for furniture, 4 for wood), then these notations can be strung together, according to the appropriate formula, with punctuation symbols dictated by the classification.

Library of Congress Classification

Main Class Outline

A	General Works
B	Philosophy. Psychology. Religion
C	Auxiliary Sciences Of History
D	World History
E	History Of The Americas
F	History Of The Americas
G	Geography. Anthropology. Recreation
H	Social Sciences
J	Political Science
K	Law
L	Education
M	Music And Books On Music
N	Fine Arts
P	Language And Literature
Q	Science
R	Medicine
S	Agriculture
T	Technology
U	Military Science
V	Naval Science
Z	Bibliography. Library Science. Information Resources (General)

Figure 11.3. Main Classes in the *Library of Congress Classification* (LCC). (Source: Library of Congress)

Both CC and UDC use punctuation marks and symbols between the notations as *facet indicators.* These help to identify the relationships among the facets. In the earliest versions of the scheme, CC used the colon (:) in significant positions; thus the name of the scheme. By the middle of the twentieth century, Ranganathan realized that a more elaborate system of facet indicators was needed to avoid confusion, and additional pieces of punctuation were adopted for this purpose. The typically long and complex notations created in faceted schemes are not easy to use for arranging physical information resources on shelves. However, in online retrieval systems they have the potential to be quite helpful, as each facet may be searched independently. Some faceted schemes, such as BC2, do not use facet indicators; instead the notations may be blended together, obscuring the individual parts.

TX703-726.3	Cookbooks
TX703-713	Early to 1800
TX714-725	1800-
TX714	General recipe collections
	Class here works consisting of collected recipes not primarily of a regional, ethnic, or international nature, nor using a specific ingredient or method of cooking, in which the technique of cooking is not emphasized
	For treatises *see* TX651
	For ethnic or regional cooking *see* TX714-725
	For international cooking *see* TX725.A1
	American
	For French, German, etc., cookbooks published in America, *see* TX719, TX721, etc.
TX715	General works
TX715.2.A-Z	By style of cooking, A-Z
TX715.2.A47	African American cooking
TX715.2.C34	California style
	Hawaiian style *see* TX724.5.H3
TX715.2.L68	Louisiana style
TX715.2.M53	Midwestern style
TX715.2.N48	New England style
TX715.2.P32	Pacific Northwest style
	Pennsylvania Dutch style *see* TX721
TX715.2.S68	Southern style
TX715.2.S69	Southwestern style
TX715.2.W47	Western style
TX715.6	Canadian
TX715.8	Greenlandic
	Latin American
TX716.A1	General works
TX716.A3-Z	By region or country, A-Z
TX717	English
TX717.2	Celtic
...	[... other regions of the world]
	Other
TX725.A1	International
TX725.A3-Z	By region or country, A-Z
	Including Africa, Australia, Middle East, etc.
TX725.M628	Middle East
(TX725.N36)	Near East *see* TX725.M628
	Cooking using alcoholic beverages
TX726	Wine. Liquors
TX726.2	Cider
TX726.3	Beer

Figure 11.4. An Illustration of Hierarchical and Enumerative Arrangement Found in LCC. (Source: Library of Congress, *Classification Web*)

LCC has limited faceting capabilities. There are tables of geographic areas in some schedules, for example, that can be used to build notations that show place. Many of the table numbers, however, are literally mathematically added to a number in the schedule; the resulting notation does not show the geographic area as a separate facet. Only the geographic area notations that are *cutter numbers*, consisting of the first letter of the geographic area name and one or more numerical digits, are recognizable as facets.

DDC has many faceting capabilities. There are six tables that can be used at various places in the classification. The notations from the tables are appended to the end of the notation from the schedules so that the facet is intact. In some cases, the end of the schedule notation and the beginning of the table notation are not demarcated; but often, with the most common subdivisions, the table notation is preceded by the digit 0. For information about DDC and LCC *call numbers* (combinations of classification notations and cutter numbers [i.e., specific notations that usually represent the primary creator, the first named creator, or the title]), see Appendix B.

Major Bibliographic Classification Schemes

The major bibliographic classification schemes (i.e., DDC, LCC, UDC) have certain components in common. These have been outlined by Hope Olson and John Boll as follows:

- A verbal description, topic by topic, of the things and concepts that can be represented in or by the scheme.
- An arrangement of these verbal descriptions in classed or logical order that is intended to permit a meaningful arrangement of topics and that will be convenient to users.
- A notation that appears alongside each verbal description, which is used to represent it and which shows the order. The entire group of verbal descriptions and notations form the schedules.
- References within the schedules to guide the classifier and the searcher to different aspects of a desired topic or to other related topics (like the related term references in alphabetical lists).
- An alphabetical index of terms used in the schedules, and of synonyms of those terms, that leads to the notations.
- Instructions for use. General instructions (with examples) are usually to be found at the beginning of the scheme, and instructions relating to particular parts of the schedules are, or should be, given in the parts to which they relate.
- An organization that will ensure that the classification scheme is maintained, that is, revised and republished. This is external to the scheme, but an important factor in evaluating its comparative usefulness.[19]

In addition to the schedules mentioned in the third point above, the schemes include tables, which are independent listings of notations for concepts that can be appended or added to the notations from the schedules. The tables comprise listings of such concepts as forms/genres, geographic areas, chronological periods, and languages. In addition to the instructions in the schemes themselves, these major bibliographic classifications have separate manuals that give further instructions and examples. Manuals can be written by the administrative organizations responsible for the maintenance of the schemes, by independent parties, or by both.

Traugott Koch has suggested that there are four broad varieties of classification schemes:[20]

- **Universal schemes**—schemes that cover the universe of knowledge (e.g., DDC, UDC, LCC, CC, BC2);
- **National general schemes**—similar to universal schemes, but designed for use in a single country (e.g., *Nederlandse Basisclassificatie* in the Netherlands and the *Sveriges Allmänna Biblioteksförening* in Sweden);
- **Subject specific schemes**—schemes that cover a particular subject area or knowledge domain (e.g., *Iconclass* for art resources; *National Library of Medicine* C*lassification*, and the *Mathematics Subject Classification* for their obvious subject matter);
- **Home-grown schemes**—schemes created to address the needs of a particular service or institution (e.g., *Metis: Library Classification for Children*, a set of 26 broad categories labeled A to Z established by four librarians who worked at the Ethical Culture School in New York City).[21]

In addition, there are a number of homegrown schemes in print that are interesting and useful in particular contexts. One example is *A Classification System for Libraries of Judaica,* now in its third edition.[22] It is kept up to date by people who use it. For example, after the publication of the third edition, a synagogue librarian created a Holocaust expansion that is posted online.[23]

Many classifications have not been mentioned in this chapter, as we have focused only on the major bibliographic classification schemes. More has been written about each of the classifications mentioned, as well as others not mentioned. Suggestions for further reading about various classification schemes (and, in some cases, manuals for applying them) are found at the end of this chapter.

CLASSIFICATION CONCEPTS

A number of classification concepts and issues affect the use of classification schemes. Some of these apply to the schemes regardless of how they are going to be used (e.g., as a way of arranging physical resources, as a way of identifying subject content in metadata, as a way of organizing and/or presenting online resources). Others are issues particularly in the use of classification as a device for arranging physical resources. The following sections are arranged from general to specific.

Broad versus Close Classification

Broad classification is classification that uses only the main classes and divisions of a scheme and perhaps only one or two levels of subdivisions. *Close classification* is classification that uses all the minute subdivisions that are available for very specific subjects. This is somewhat like the specific versus general issue in controlled vocabularies. When beginning to use a classification scheme it is necessary to decide whether to use only the top levels of the scheme or whether to use the scheme at the deepest level possible (or something in between).

One issue here is that if the intent of using the scheme is to collocate topics, then broad versus close may depend upon the size of the collection that is being classified. If the collection is very large, then using only the top levels of the scheme means that very large numbers of resources, or the surrogate records for them, will be collocated at the same notation. On the other hand, if the collection is very small, then using close classification may mean that most notations are assigned to only one or two resources, with the result that collocation is minimal, unless one is using a scheme like DDC where one can drop digits off the end of the notation to get to the next broader level of the concept being sought.

It should be noted that what is considered close classification for a small collection may be broad classification for a large collection. That is, for a small collection, DDC's 612.1, meaning the concept of *blood and circulation* in medicine, is a very specific level; but in a medical collection there may be a thousand or more information resources on blood and circulation, and for that collection, close classification would mean going to, for example, 612.1127 for *white corpuscle counts.*

Another issue has to do with the globalization of information organization. Even if a collection in a particular place is small, its metadata may be combined with that of many collections, producing a mega-collection. If some of the metadata has been created using close classification, and some has been created using broad classification, the combined effect will be confusing and less helpful. It may be necessary in today's world to use the closest classification available in the scheme being used. This is problematic in some institutions where the classification is used for arranging physical resources. Close classification often produces very long notations, which sometimes have to be placed on very small resources. Michael Gorman has suggested that this problem be solved by using shorter notations for *call numbers* but using close classification for the purposes of intellectual retrieval.[24] (See discussion of call numbers in Appendix B.) This, of course, is objected to in many quarters because of the extra cost of providing two notations; but there is a cost to the user who is trying to retrieve information when classification is not usable for collocating topics.

Classification of Knowledge versus Classification of a Particular Collection

Classification of knowledge is the concept that a classification system can be created that will encompass all knowledge that exists. DDC began as a classification of knowledge—at least Western knowledge as understood by Melvil Dewey. Classification of a particular collection is the concept that a classification system should only be devised for the information resources that are being added to collections, using the concept of *literary warrant.* Literary warrant is the name applied to the concept that new notations are created for a classification scheme and new terms are added to a controlled vocabulary only when information resources actually exist that are about a new concept. (See also comments on literary warrant in Chapter 10.) LCC began as a classification of a particular collection.

Even though DDC began as a classification of knowledge, it has been forced to use literary warrant for updates and revisions. Some areas of knowledge that have developed in the twentieth century need significantly more space in the scheme than originally given, if indeed they existed and were given any space at all (e.g., computer science). Dewey devoted a whole division to the artificial waterways called *canals,* but the concept has been moved to allow expansion of other areas such as engineering; canals are no longer being given the attention they were given in Dewey's day. And this is one value of the approach taken by LCC. The alphabet has 26 letters. In most places in LCC letters are doubled, giving the potential for up to 676 divisions versus the 100 available to DDC (and if three letters are used, LCC could offer up to 17,576 sections vs. DDCs 1,000 sections!). LCC has not come close to using this amount of space.

It can be seen that if a scheme has been devised using literary warrant, it can be more flexible than one devised on the basis of classification of knowledge. This issue may come down to preference as to the logical approach of a classification of knowledge versus the practical approach of a classification of what is being studied, created, or written about at the present time.

Integrity of Numbers versus Keeping Pace with Knowledge

Integrity of numbers is the concept that in the creation and maintenance of a classification scheme, a notation, once assigned, should always retain the same meaning and should never be used with another meaning. *Keeping pace with knowledge* is the concept that in the creation and maintenance of a classification scheme, it is recognized that knowledge changes, and therefore it is necessary to be able to move concepts, insert new concepts, and to change meanings of numbers.

Melvil Dewey was a strong advocate of the integrity of numbers. He did not want the users of his system ever to have to change a number on a resource because the number's meaning had been changed in the classification. He wanted new concepts to be assigned to new numbers. As the twentieth century went forward, however, it became impossible to keep pace with new knowledge without sometimes changing the older notations. For example, in the field of mathematics the understanding of the field changed with new research. So it became necessary to change Dewey's arrangement of the basic sections of mathematics, with the following result:

DDC 1st edition	DDC 23rd edition
510 Mathematics	**510 Mathematics**
511 Arithmetic	**511 General principles of mathematics**
512 Algebra	**512 Algebra**
513 Geometry	**513 Arithmetic**
514 Trigonometry	**514 Topology**
515 Conic sections	**515 Analysis**
516 Analytical geometry	**516 Geometry**
517 Calculus	**517** ***Unassigned***
518 Quaternions	**518 Numerical analysis**
519 Probabilities	**519 Probabilities and applied mathematics**

At the local level, reassignment of numbers always involves soul searching on the part of catalogers. As cost is an ever-present issue, there is a desire not to have to change notations because this can be expensive. On the other hand, if changes are not made, then the digits 513 would mean both *geometry* and *arithmetic,* and consequently, collocation would be compromised. In such a scenario, both searching for classification in the catalog and browsing the shelves could confuse the user. It has been argued that leaving the older materials in the older numbers accurately reflects their contents, while newer materials reflect the new thinking that prompted the change in the scheme. This only works if the different schemes are kept separate both in arranging and in searching.

At the level of the classification scheme itself, notation and structural changes also involve soul searching on the part of those designing and maintaining the classification. They worry about the impact of changes on those applying the system, but they must update the classification periodically or it will become irrelevant and out of date, and consequently the scheme will not continue to serve its purpose. If it is flexible enough, updates can be accomplished by inserting new notations. LCC, for example, accomplishes most of its updating in this fashion. If the scheme is less flexible, as with DDC, some inserting can be done, but sometimes the meanings of numbers must also be changed.

It was mentioned earlier that some classification issues are general, but others are issues for the use of classification for arranging physical resources. *Integrity of numbers* versus *keeping pace with knowledge* is an issue that has both general and physical implications and gives us a transition into the issues regarding physical entities. In this case, reclassification of physical resources when meanings of numbers have been changed is an expensive process. In most collections complete reclassification is not done. Instead there often is some kind of process set up to reclassify resources as they are used. Resources needing reclassification are taken care of as they are returned after the first use after a number's meaning has changed. Or sometimes, all of the resources affected by a change of numbers in a whole section of the scheme are reclassified on a project basis. Changing only some of a collection, of course, ignores the use of classification as a search key in an online system. Searching a system where certain notations bring up surrogate records for information resources on different subjects is not satisfactory.

Closed versus Open Stacks

Closed stacks is the name given to the situation where information resource storage areas are accessible only to the staff of the library, archives, or other place that houses information resources. *Open stacks* is the name given to the opposite situation, where patrons of the facility have the right to go into the storage areas themselves. In closed stack situations users must call for resources at a desk and then wait for them to be retrieved and delivered. This eliminates any possibility of what is called *browsing* in the stacks. Browsing is a process of looking, usually based on subject, at all the resources in a particular area of the stacks, or in a listing in an online retrieval tool, in order to find, often by serendipity, the resources that best suit the needs of those browsing. In a closed stacks situation one can browse only in the catalog—not always an easy process, depending upon the system design.

In cases where remote storage is used for older and less-used materials, compact shelving is often used. This means that in order to make the most efficient use of space, resources are shelved using every inch of shelf space, usually with fixed locations. In such situations the storage areas are closed. A number of large research libraries have had closed stacks for a long time. There are various reasons for this, including tradition, vandalism, and precedence of certain classes of users (e.g., faculty, graduate students) over others. In such libraries, a proposal to stop classifying resurfaces every so often. The question asked is, "Why classify if readers cannot browse?" Such a proposal includes data about how costs will be lowered if classification is stopped. However, classification is a major form of subject access, and if browsing of the stacks is not allowed, then browsing of the classification listing in the catalog becomes even more important. Reference librarians often use classification to assist users in finding subject-related material.[25]

In archives, the storage areas are almost always closed. This hardly matters as far as classification goes because classification has not been found to be particularly useful for archives in any case. A collection of records can have individual pieces that are on diverse subjects, and dividing and separating these out to classify would violate the principle of *respect des fonds*. It is conceivable that whole collections could be classified, but usually the classification in such a case would be so broad as to be nearly meaningless.

Fixed versus Relative Location

The term *fixed location* signifies a set place where a physical information resource will always be found or to which it will be returned after having been removed for use. A fixed location identifier can be an accession number; or a designation made up of room number, stack number, shelf number, position on shelf; or other such designations. The term *relative location* is used to mean that an information resource will be or might be in a different place each time it is reshelved; that is, it is reshelved relative to what else has been acquired, taken out, returned, and so forth, while it was out for use. The method for accomplishing this is usually a *call number* with the top line or two being a classification notation. (See discussion of call numbers in Appendix B.)

Fixed location is often used for the purpose of saving space. If fixed location is used, space does not have to be left at the end of each shelf for potential new acquisitions in a particular area of the classification. This is particularly useful for remote storage facilities where space must be saved. Relative location, however, is desirable in a situation where users have access to the stacks and can browse (see discussion of closed versus open stacks above). It is often argued that the cost of fixed location is much less than that of relative location. If classification notations do not have to be assigned, and if space can be used to the fullest, then the cost of fixed location is less to the agency involved. However, especially in libraries, the cost is passed on to the user who loses the collocation provided by classification and the serendipity of being able to browse in the stacks.

Location Device versus Collocation Device

Another controversy surrounding classification is that of whether classification serves as a collocation device or whether it is simply a location device. A *location device* is a number or other designation on a resource to tell where it is located physically. It can be, among other designators, an accession number, a physical location number, or a call number. A *collocation device* is a number or other designation on a resource used to place it next to (i.e., collocate with) other resources that are like it. It is usually a classification notation.

Those responsible for the costs of organizing collections tend to take the location device view. They often believe that any call number is fine as long as the notation placed on the physical resource matches the notation in the metadata for the resource. The argument is that if the notations match, the user will be able to find the resource. However, this assumes that the subject headings are sufficient for finding subject-related material in a catalog. Thomas Mann, cited above, and others have shown that both subject headings and classification are required for the most effective retrieval of subject-related material.[26]

It is not clear whether the same notation is adequate to serve both the collocation and the location functions. Gorman, as mentioned above, has suggested that a fully detailed classification be assigned for the purpose of collocation, while a shortened version of it be used for a location device.[27] (This would not really be helpful in the case of LCC, because most notations, even for complex subjects, are relatively short.) In most of the United States, one classification notation has served for both collocation and location for many decades. In some other places, especially where classified catalogs are used, the functions of collocation and location have been served by different notations for decades.

A recent point for discussion has been the classification of online resources. Cost managers often believe that classification is mainly for location, and because online resources have location devices in the form of URLs, no classification is needed. However, for more complete collocation of like resources (no matter the format), classification can be an effective means of organizing online information resources for subject retrieval.

Classification of Serials versus Alphabetical Order of Serials

Serials are sometimes called *journals* or *magazines.* A *serial* may be defined as a publication issued in successive parts (regularly or irregularly) that is intended to continue indefinitely. Classification of serials means that a classification notation is assigned to a serial, and this classification notation is placed on each bound volume and/or each issue of that serial. Alphabetical order of serials means that the serials are placed in order on shelves or in a listing according to the alphabetical order of the titles of the serials. The concerns involved in the classification versus alphabetical order dilemma apply to runs of printed serials. Serials produced only online bring us questions and problems, but classification is not one of these.

With serials it is sometimes quite difficult to discern the intended title because of its placement (e.g., following a corporate body name that is in larger type than the title). In addition, serials are apt to change titles, sometimes several times. They also can merge or split, and can be absorbed by another serial, all of which result in title changes. This is one of the most persuasive arguments against alphabetical order of serials. What to do with runs of serials when there has been a title change is not a resolved issue. Some institutions move the whole run and place, at the alphabetical position of the old title(s), a block or sign that provides a cross-reference that gives the new title. It is argued in this situation that users will expect to find all of that serial together in spite of the title change. Other institutions divide the serial at the point of the new title. It is argued here that when a title changes, it does so because of a change of emphasis in the serial; therefore, the two runs are almost like two different serials. In addition, a user who goes into the alphabetically arranged collection with a citation may only know the title that was on the serial at the time the article being sought was published.

Classification of serials provides a solution to the title change dilemma. The new title can be given the same classification as the old. It also has the advantage of allowing the serial to be placed with monographs on the same subject; this is beneficial for browsing purposes. However, classification requires two look-ups. That is, a user comes to the collection with a citation to an article in a journal. If the serial collection is classified, the user must look for the journal title in the catalog, note its call number, and then go to the stacks to find that call number. If serials are placed alphabetically, the user goes to the serials stacks with the citation and finds the journal title. Classification also requires more cataloging/indexing time and thus seems to cost more, although the maintenance of title changes in an alphabetically arranged situation may offset some of the cost that would be incurred by classifying.

One other consideration is the arrangement of serials in a special library. Many such libraries are subject area specific. In some of these situations, most of the serials are on the same or closely related subjects, which could mean that their classification notations would be very near each other, resulting in a semi-alphabetical order. Such libraries often choose alphabetical arrangement in the first place. The actual arrangement of serials in all libraries in the United States is about 50 percent alphabetical and 50 percent classified.

Classification of Monographic Series (Classified Separately versus Classified as a Set)

A *monograph* is a complete bibliographic unit or information resource. It is often a single work but may also be one work or more than one work issued in successive parts; but unlike a serial, it is not intended to be continued indefinitely. A work intended to be complete in, let's say, 28 volumes, is a monograph. A *monographic series* is a hybrid of monograph and serial. In a monographic series each work that is issued as a part of the series is a monograph but is identified as one work in the series. The series itself may or may not be intended to be continued indefinitely. For example, the *Library and Information Science Text Series* published by Libraries Unlimited continues to have new works added to it as appropriate works are identified by the publisher.

The difficulty with classification of monographic series is whether to classify each work in the series separately with a specific notation representing its particular subject matter, or to treat the series as if it were a multivolume monograph and give all parts of the series the same (usually broad) classification notation representing the subject matter of the entire series. Classification of the series as a set results in the loss of collocation of the specific subjects contained in each of the works.

On the other hand, especially in public libraries, classification of the series as a set may meet with resounding approval by the users. There are readers who wish to go through a series one by one and find it satisfactory to have them classified together. It is obviously the intention of some publishers that some series be shelved together, especially those series that are bound alike. In some cases the situation is handled by acquiring duplicates and then classifying one with the set and the other in its specific subject area. The cost of doing this usually is not supported, however, and sometimes cost is the deciding factor for classifying a series as a set; it is cheaper to slap on an already-conceived classification notation than to do the subject analysis necessary for specific classification.

THE USE OF CATEGORIES AND TAXONOMIES ONLINE

The use of categories or taxonomies (the terminology seems to depend upon the site) is readily apparent online. On a commercial shopping site like Amazon.com[28] or a website for a large-scale retailer, such as the grocery store known as Wegmans,[29] a taxonomy is often found directly on the homepage or in a pull-down menu (or both). An information architect created this taxonomy to provide shoppers with a quick, user-friendly set of links to the main divisions in order to improve navigation and the overall experience of the site, as well as the ability to find information and products quickly and efficiently; the following are the top-level categories provided in Wegmans' taxonomy:

- Baby Essentials
- Bakery
- Beer Shop
- Beverages
- Bulk Foods
- Cheese Shop
- Coffee Shop

- Dairy
- Deli
- Floral
- Frozen Foods
- Grocery Food
- Health and Wellness
- Home and Entertaining
- Household Essentials
- International Foods
- Kosher
- Meat
- Nature's Marketplace
- Packaged Meals, Entrees, and Sides
- Personal Care and Makeup
- Pet Supplies
- Pizza
- Prepared Foods–Market Cafe
- Produce
- Seafood
- Sub Shop
- Sushi
- Wine & Spirits

A second level of hierarchy contains an additional set of headings. Taking the heading *bakery* as an example, the second level offers the following categories to explore:

- Bread, Fresh Baked
- Bread, Packaged
- Breakfast
- Celebration cakes
- Cookies
- Desserts
- Miscellaneous
- Pitas, Flatbreads, & Wraps
- Rolls
- Special Diet
- Stuffing

Under the category *desserts*, a third layer of hierarchy is present. These sub-categories include:

- Bar
- Breakfast Buns
- Brownies

- Cake
- Candy-Patisserie
- Cannoli
- Cheesecake
- Cookies
- Cupcakes
- Dessert
- Eclairs
- Mousse Cups
- Pie
- Puff
- Shortcake
- Tart
- Tiramisu
- Trifles

In addition to these sub-categories, a Wegmans-specific search engine is also provided at each level, and additional facets are provided to narrow further the results (e.g., brands, 5-star ratings, whether the product is vegan or not). The taxonomy offers users, who may not know what they want, an option to browse, exploring the various products available in the commercial venture's inventory. The difficulty with taxonomies at some commercial sites like Amazon.com is the overwhelming size of the overall inventory, and the unpredictability of the hierarchies employed. Even at the deepest, most specific levels of a taxonomy, there still may be too many resources to browse and the user still may need to conduct a keyword search.

Taxonomies also are used by noncommercial sites, such as the National Center for Biotechnology Information's public databases.[30] But, perhaps, the most well-known example of using categorization online was the Yahoo! Directory, where, before it closed in 2014, websites were placed into categories created by Yahoo! indexers, and the categories were browsed in hierarchical fashion. The Yahoo! Directory is no longer in use, but there are some sites that still use this approach to organizing lists of Internet resources. Two examples are the Best of the Web directory[31] and the WWW Virtual Library.[32]

As was seen at Wegmans, sub-categories in "Best of the Web" are hierarchically nested under these top-level categories:

- Arts
- Business
- Computers
- Finance
- Games
- Health
- Home
- Kids & Teens
- News
- Recreation
- Reference
- Science
- Shopping
- Society

- Sports
- Regional

The hierarchies, however, are not true hierarchies in the sense that they represent actual genus–species relationships. In other words, the categories at the second hierarchical level are not necessarily a type or kind of the broader concept at the first level, nor are the sub-categories at the second level necessarily at equivalent levels of specificity with each other. For example, under *Society* one finds 37 sub-categories (the boldface terms are categories also located in other parts of the taxonomy, as explained below):

- Activism
- Advice
- Blogs
- Crime
- Death
- Disabled
- **Economics**
- **Education**
- Ethnicity
- Folklore
- Future
- Gay, Lesbian, and Bisexual
- Genealogy
- Government
- History
- Holidays
- Issues
- **Language and Linguistics**
- Law
- Lifestyle Choices
- **Men**
- Military
- Organizations
- Paranormal
- People
- Philanthropy
- Philosophy
- Politics
- Relationships
- Religion
- Sexuality
- **Social Sciences**
- **Sociology**
- Subcultures

- Transgendered
- **Women**
- Work

In examining the terms in the list, one can see a strange mix of concepts. It is unusual to see categories like *Economics* and *Education* alongside *Men* and *Holidays*. Several sub-categories are obviously related to others in the top level of the hierarchy. For example, the terms highlighted with boldface type are categories that are primarily located in other parts of the taxonomy. *Economics* and *Language and Linguistics* are actually placed under *Social Sciences*, which is listed as a sub-category under both *Science* and *Society*. *Men* and *Women* are sub-categories under *People*; they are both under and collocated alongside *People*. Categories such as *Activism*, *Death,* and *Transgendered* are not at the same level of specificity as *History*. This is obviously not a strict hierarchical structure.

CATEGORIZING SEARCH RESULTS

Some researchers believe that search satisfaction and effectiveness could be improved if classification techniques were applied to search engines.[33] The use of classification in conjunction with search engines is not a new idea. In the late 1990s and early 2000s, the now defunct Northern Light, a free and publicly available search engine, used document-clustering techniques to divide search results into broad subject categories. While it did not work directly with a formal classification system, Northern Light used document clustering to improve the user's ability to navigate through results sets. At the time, Northern Light was the only major search engine to use these techniques for displaying results. By 2001 Northern Light had stopped offering a public web search engine, instead focusing on commercial enterprises.[34] Since then, research has continued on automatic classification techniques, and other clustering search engines have been developed. For example, Carrot[35] offers several approaches to clustering results, some visual and some text based, and Yippy[36] provides individual page results along with a list of categories to help users sharpen or focus their search results.

All of these projects use document-clustering techniques to achieve these navigable, user-friendly results sets. What is *document clustering*? According to Srividhya Subramanian and Keith Shafer, clustering is a computer science approach to classifying documents in a collection, based on the contents of the entire collection. They state, "It explores collection relationships by computing the similarity between every pair of documents in the collection. Using the similarity information, clustering attempts to divide a collection of documents into groups (clusters)."[37] As a result, documents in the same clusters are more similar than documents in different clusters. Because these techniques are automated, they are not without errors (e.g., some odd pairings can occur), but overall, the technology is becoming more sophisticated and can provide impressive results.

CONCLUSION

This chapter has addressed categorization and classification as a way to provide subject access to information resources. One of the most important factors in information retrieval is to have resources and metadata records arranged and displayed in a logical fashion. Classification often plays a major role in this arrangement and display. Familiarity with the history of categories and the basics of classification theory can be helpful in understanding and using the classification schemes employed today. Arrangement of classification schemes (i.e., hierarchical, enumerative, and faceted) is based in classification theory. Specific classifications are numerous and run from universal to homegrown schemes. Regardless of which scheme is used, there are classification concepts and controversies that must be addressed in the application of any scheme. Some concepts apply to physical resources, while others apply both to physical and intangible resources. A number of online sites that organize resources or products are using categories, classifications (often called taxonomies in this environment), or clustering techniques. Automatic techniques are growing in number and in sophistication.

Some Important Terms in This Chapter

Analytico-Synthetic Classification	Closed Stacks	Literary Warrant
Broad Classification	Collocation Device	Location Device
Browsing	Cutter Number	Monograph
Call Number	Document Clustering	Monographic Series
Categorization	Enumerative Classification	Open Stacks
Category	Facet	PMEST
Class	Facet Indicators	Relative Location
Classification	Faceted Classification	*Respect des Fonds*
Classification Scheme	Fixed Location	Schedules
Close Classification	Hierarchical Classification	Serials
		Taxonomy

Some Important Acronyms in This Chapter

BC: *Bibliographic Classification*
BC2: *Bibliographic Classification*, 2nd ed.
CC: *Colon Classification*
CRG: Classification Research Group
DDC: *Dewey Decimal Classification*
EC: *Expansive Classification*
LIS: Library and Information Science
LCC: *Library of Congress Classification*
NLM: National Library of Medicine
OED: *Oxford English Dictionary*
UDC: *Universal Decimal Classification*

NOTES

All URLs accessed June 2017.

1. Oxford English Dictionary Online, available by subscription only, http://dictionary.oed.com.

2. Merriam-Webster Online Search, http://www.merriam-webster.com/.

3. Elin K. Jacob, "Classification and Categorization: A Difference That Makes a Difference," *Library Trends* 52, no. 3 (Winter 2004): 522.

4. Ibid., 527.

5. P. A. Studer, "Classification as a General Systems Construct," in *Information Management in the 1980's: Proceedings of the 40th ASIS Annual Meeting, Chicago, Illinois, September 26–October 1, 1977,* eds. B. M. Fry and C. A. Shepherd (White Plains, NY: Knowledge Industry for American Society for Information Science, 1977), 67 and microfiche nos. 8–9, G6–A9.

6. Christopher J. Dent, "Old Stuff: Classification v Categorization," in "Glacial Erratics," http://www.burningchrome.com/~cdent/mt/archives/000401.html.

7. Ibid.

8. Amy J. Warner, "A Taxonomy Primer," http://www.ischool.utexas.edu/~i385e/readings/Warner-aTaxonomyPrimer.html.

9. “Ten Taxonomy Myths,” *Montague Institute Review* (November 2002; updated April 29, 2003), https://web.archive.org/web/20141216163734/http://www.montague.com:80/review/myths.html.

10. Thomas D. Wason, “Dr. Tom’s Taxonomy Guide,” http://www.tomwason.com/drtomtaxonomiesguide.html.

11. F. W. Lancaster, *Indexing and Abstracting in Theory and Practice,* 3rd ed. (Champaign: University of Illinois, Graduate School of Library and Information Science, 2003), xiii.

12. Bertram C. Bruce and James A. Levin, “Educational Technology: Media for Inquiry, Communication, Construction, and Expression,” *Journal of Educational Computing Research* 17, no. 1 (1997): 79–102, https://www.ideals.illinois.edu/bitstream/handle/2142/13421/media_inquiry_communication_construction_xp sequence=2&isAllowed=y.

13. U.S. Dept. of Agriculture, Agricultural Research Service, “GRIN Taxonomy for Plants,” https://npgsweb.ars-grin.gov/gringlobal/taxon/taxonomyquery.aspx.

14. Douglas Harper, “Online Etymology Dictionary,” http://www.etymonline.com/index.php?term=category.

15. Aristotle, *Categories,* trans. E. M. Edghill (Adelaide: eBooks@Adelaide, University of Adelaide Library, 2007), section 4, https://ebooks.adelaide.edu.au/a/aristotle/categories/ and http://www.classicallibrary.org/aristotle/categories/index.htm.

16. George Lakoff, *Women, Fire, and Dangerous Things: What Categories Reveal about the Mind* (Chicago: University of Chicago Press, 1987), 6.

17. Ibid., 16–57.

18. Stephen L. Zegura, “Genes, Opsins, Neurons, and Color Categories: Closing the Gaps,” in *Color Categories in Thought and Language,* eds. C. L. Hardin and Luisa Maffi (Cambridge: Cambridge University Press, 1997), 289.

19. Hope A. Olson and John J. Boll, *Subject Analysis in Online Catalogs,* 2nd ed. (Englewood, CO: Libraries Unlimited, 2001), 155–56.

20. Traugott Koch and Michael Day, “The Role of Classification Schemes in Internet Resource Description and Discovery: Executive Summary,” http://www.ukoln.ac.uk/metadata/desire/classification/class_su.htm.

21. Metis, https://sites.google.com/site/metisinnovations/.

22. David H. Elazar and Daniel J. Elazar, *A Classification System for Libraries of Judaica,* 3rd ed., with the assistance of Rachel K. Glasser and Rita C. Frischer (Northvale, NJ: Jason Aronson, 1997).

23. Carylyn Gwyn Moser, “Elazar Classification System Holocaust Expansion,” https://sites.google.com/site/mtevansco/elazar-classification.

24. Michael Gorman, “The Longer the Number, the Smaller the Spine; or, Up and Down with Melvil and Elsie,” *American Libraries* 12, no. 8 (September 1981): 498–99.

25. Thomas Mann, *The Oxford Guide to Library Research,* 4th ed. (Oxford: Oxford University Press, 2015), Chapter 3.

26. Ibid.

27. Gorman, “The Longer the Number,” 498–99.

28. Amazon.com, http://www.amazon.com/.

29. Wegmans, https://www.wegmans.com/.

30. National Center for Biotechnology Information, “Taxonomy,” https://www.ncbi.nlm.nih.gov/taxonomy.

31. Best of the Web Directory, https://botw.org/.

32. The WWW Virtual Library, http://www.vlib.org/.

33. Young Mee Chung and Young-Hee Noh, "Developing a Specialized Directory System by Automatically Classifying Web Documents," *Journal of Information Science* 29, no. 2 (2003): 118.

34. Northern Light, http://www.northernlight.com.

35. Carrot2, http://search.carrot2.org/stable/search.

36. Yippy, http://yippy.com/.

37. Srividhya Subramanian and Keith Shafer, "Clustering," *Journal of Library Administration* 34, no. 3/4 (2001): 221–28.

SUGGESTED READINGS

Classification in General

Bowker, Geoffrey C., and Susan Leigh Star. *Sorting Things Out: Classification and Its Consequences.* Cambridge, MA: MIT Press, 1999. Introduction: "To Classify Is Human," Chapter 9: "Categorical Work and Boundary Infrastructures: Enriching Theories of Classification," and Chapter 10: "Why Classifications Matter."

Broughton, Vanda. *Essential Classification*. Chicago: ALA Neal-Schuman, 2015.

Chan, Lois Mai, and Athena Salaba. *Cataloging and Classification: An Introduction*. 4th ed. Lanham, MD: Rowman & Littlefield, 2016. Chapter 15: "Classification and Categorization."

Classification Research Group. "The Need for a Faceted Classification as the Basis of All Methods of Information Retrieval." In *Theory of Subject -Analysis: A Sourcebook,* edited by Lois Mai Chan, Phyllis A. Richmond, and Elaine Svenonius, 154–67. Littleton, CO: Libraries Unlimited, 1985.

Gorman, Michael. "The Longer the Number, the Smaller the Spine; or, Up and Down with Melvil and Elsie." *American Libraries* 12, no. 8 (September 1981): 498–99.

Hunter, Eric J. *Classification Made Simple: An Introduction to Knowledge Organisation and Information Retrieval.* 3rd ed. Farnham, UK: Ashgate, 2009.

Joudrey, Daniel N., Arlene G. Taylor, and David P. Miller. *Introduction to Cataloging and Classification*. 11th ed. Santa Barbara, CA: Libraries Unlimited, 2015. Chapter 16: "Classification."

Lakoff, George. *Women, Fire, and Dangerous Things: What Categories Reveal about the Mind.* Chicago: University of Chicago Press, 1987. Chapter 1: "The Importance of Categorization."

Langridge, D. W. *Classification: Its Kinds, Elements, Systems, and Applications.* London: Bowker-Saur, 1992.

Leise, Fred. "Using Faceted Classification to Assist Indexing." *Key Words* 9, no. 6 (November/December 2001): 178–79.

Mann, Thomas. *The Oxford Guide to Library Research.* 4th ed. Oxford: Oxford University Press, 2015. Chapter 3: "General Browsing, Focused Browsing, and Use of Classified Bookstacks."

Marcella, Rita, and Arthur Maltby, eds. *The Future of Classification.* Aldershot, UK: Gower, 2000. Chapter 1: Eric Hunter. "Do We Still Need Classification?"

Olson, Hope A., and John J. Boll. *Subject Analysis in Online Catalogs.* 2nd ed. Englewood, CO: Libraries Unlimited, 2001. Chapter 7: "Bibliographic Classification."

Soergel, Dagobert. "The Rise of Ontologies or the Reinvention of Classification." *Journal of the American Society for Information Science* 50, no. 12 (October 1999): 1119–20.

Specific Classification Schemes

Dewey Decimal Classification

Foskett, A. C. *The Subject Approach to Information.* 5th ed. London: Library Association Publishing, 1996. Chapter 17: "The Dewey Decimal Classification."

Joudrey, Daniel N., Arlene G. Taylor, and David P. Miller. *Introduction to Cataloging and Classification*. 11th ed. Santa Barbara, CA: Libraries Unlimited, 2015. Chapter 17: "Decimal Classification."

OCLC. "Dewey Services." http://www.oclc.org/en/dewey.html.

Olson, Hope A., and John J. Boll. *Subject Analysis in Online Catalogs.* 2nd ed. Englewood, CO: Libraries Unlimited, 2001. Chapter 8: "Online Catalogs and the Dewey Decimal and Library of Congress Classifications."

Universal Decimal Classification

Foskett, A. C. *The Subject Approach to Information.* 5th ed. London: Library Association Publishing, 1996. Chapter 18: "The Universal Decimal Classification."

Joudrey, Daniel N., Arlene G. Taylor, and David P. Miller. *Introduction to Cataloging and Classification*. 11th ed. Santa Barbara, CA: Libraries Unlimited, 2015. Chapter 17: "Decimal Classification."

McIlwaine, Ia. *The Universal Decimal Classification: A Guide to Its Use*. The Hague: UDC Consortium, 2007.

UDC Consortium. "About Universal Decimal Classification and the UDC Consortium." http://www.udcc.org/about.htm.

Library of Congress Classification

Chan, Lois Mai, Sheila S. Intner, and Jean Weihs. *A Guide to the Library of Congress Classification.* 6th ed. Santa Barbara, CA: Libraries Unlimited, 2016.

Foskett, A. C. *The Subject Approach to Information.* 5th ed. London: Library Association Publishing, 1996. Chapter 22: "The Library of Congress Classification."

Joudrey, Daniel N., Arlene G. Taylor, and David P. Miller. *Introduction to Cataloging and Classification*. 11th ed. Santa Barbara, CA: Libraries Unlimited, 2015. Chapter 18: "Library of Congress Classification (LCC)."

Library of Congress. "Library of Congress Classification." https://www.loc.gov/catdir/cpso/lcc.html.

Olson, Hope A., and John J. Boll. *Subject Analysis in Online Catalogs.* 2nd ed. Englewood, CO: Libraries Unlimited, 2001. Chapter 8: "Online Catalogs and the Dewey Decimal and Library of Congress Classifications."

Snow, Karen. *A Practical Guide to Library of Congress Classification*. Lanham, MD: Rowman & Littlefield, 2017.

Colon Classification

Foskett, A. C. *The Subject Approach to Information.* 5th ed. London: Library Association Publishing, 1996. Chapter 21: "The Colon Classification."

Joudrey, Daniel N., Arlene G. Taylor, and David P. Miller. *Introduction to Cataloging and Classification*. 11th ed. Santa Barbara, CA: Libraries Unlimited, 2015. Chapter 16: "Classification."

Ranganathan, S. R. *Prolegomena to Library Classification.* 3rd ed. London: Asia Publishing House, 1967.

Steckel, Mike. "Ranganathan for IAs: An Introduction to the Thought of S.R. Ranganathan for Information Architects." *Boxes and Arrows* (October 7, 2002). http://boxesandarrows.com/ranganathan-for-ias/.

Other Classification Schemes

Foskett, A. C. *The Subject Approach to Information.* 5th ed. London: Library Association Publishing, 1996. Chapter 19: "The Bibliographic Classification" and Chapter 20: "The Broad System of Ordering."

Joudrey, Daniel N., Arlene G. Taylor, and David P. Miller. *Introduction to Cataloging and Classification*. 11th ed. Santa Barbara, CA: Libraries Unlimited, 2015. Chapter 20: "Other Classification Systems."

CONCLUSION

In the preceding chapters we have discussed the past and present of information organization. Organizing has been going on, we suspect, since the first appearance of humans, because it seems to be such an innate need. Even some nonhuman animals, birds, and insects are organizers (e.g., ants have highly organized societies with certain responsibilities assigned to various individuals). The organization of information among humans began at least as soon as language developed sufficiently for information to be passed from person to person. In oral traditions information was organized in people's minds, and, in recorded information traditions, ways of identifying information resources were developed at least as early as clay tablets.

Current retrieval tools, including bibliographies, catalogs, indexes, finding aids, museum registers, databases, and search engines, represent the highest state of progress in information organization that humans have achieved so far. These tools are far from perfect, though, and there is much room for organizers to improve metadata and for system designers to provide more helpful retrievals and more logical displays. Development of seamless interfaces among these tools, which once was seen only as a possibility, has become a necessity; more work in this area has to be done.

Arrival at this current state of organization has taken centuries. Progress in the Western world was given impetus by printing with moveable type, but even then, movement toward bibliographic control was slow as printing moved through the stage of imitating manuscript production to the stage of creating new type fonts just for printing. We are experiencing a similar stage now as we use computers to imitate our previous print culture. Alan Kay, the first person to conceive of the laptop computer, has said that the computer revolution has not yet begun:

> The printing press was invented in the middle of the 15th century, yet it took 100 years before a book was considered dangerous enough to be banned, 150 years before science was invented, almost 200 years before a new kind of political essay was invented, and more than 300 years before a country with an invented political system (the US) could be argued into existence via the press and a citizenry that could understand the arguments. Schooling and general literacy were also fruits of the press, and also took many centuries to become established. The commercial computer is now about 50 years old and is still imitating the paper culture that came before it, just as the printing press did with the manuscript culture it gradually replaced. No media revolution can be said to have happened without a general establishment of "literacy": fluent "reading" and "writing" at the highest level of ideas that the medium can represent. With computers we are so far from that fluent literacy—or even understanding what that literacy should resemble—that we could claim that the computer revolution hasn't even started.[1]

Although it seems that much has happened in the 20 years since Kay said this, we still have not accomplished the literacy of which he spoke. Therefore, we cannot know what organized information will look like in another 50 years or even 25 years. We believe, however, that the principles of organization that have developed over the last several hundred centuries will not be thrown out but will continue to evolve into the organizing principles of the future.

NOTE

1. Alan Kay, "The Computer Revolution Hasn't Happened Yet," lecture in the SCS Distinguished Lecture Series, presented at Carnegie Mellon University, April 29, 1998.

APPENDIX A: AN APPROACH TO SUBJECT ANALYSIS

A way to use the subject analysis concepts presented in Chapter 9 is to follow the outline presented below. It attempts to take the analyst through the most commonly used features of a work, to address the concepts that are most helpful in determining aboutness, and to consider some questions that might help to make the process less vague. It is followed by an example providing subject analysis of a popular website.

I. **Analyze the information resource**
   A. Identify the overall discipline, branch of knowledge, or subject area in which the work fits. Consider the intended audience and the purpose of the information resource (e.g., why the resource was created, what it accomplishes, what questions it might answer, and for whom).
   B. Identify the important topics or concepts in the information resource. Look for frequently mentioned topics or ideas in the following:
      1. Title and subtitle.
      2. Table of contents, chapter titles, section headings, or equivalents.
      3. Preface and/or introduction, first chapter, etc.
      4. Illustrations and their captions.
      5. Conclusion (e.g., last chapter, section, etc.).
   C. Identify names used as subject concepts.
   D. Identify role(s) of any geographic name(s) present.
   E. Identify chronological elements.
   F. Identify form or genre of the resource being analyzed.

II. **Construct an aboutness statement**
   A. Describe what the information resource is about.
   B. Review the resource again. Look for support for your ideas. Refine the statement as needed.
   C. Identify terms from the aboutness statement to be searched in the controlled vocabulary.
   D. Sketch out a rudimentary hierarchy (discipline/sub-discipline/concept/topic, etc.) into which the aboutness concepts fall.

III. **Translate the aboutness into controlled vocabulary**
   A. Search the chosen terms in the controlled vocabulary. Translate the terms into the specific headings or descriptors used in the controlled vocabulary list.
   B. Using the hierarchy as a guide, search the classification for the most appropriate class(es) to describe the resource's aboutness. Convert the aboutness into specific classification notation(s).
   C. Review your choices and compare to similar items.

EXAMPLE

The summarization-level analysis below is for the website *JGarden: The Japanese Garden Database* created and maintained by Robert Cheetham (please see http://www.jgarden.org). Analysis of the website and construction of subject access mechanisms might proceed as follows:

I. Analyze the information resource

A. Identify the overall discipline, branch of knowledge, or subject area into which the work fits. Consider the intended audience and the purpose of the information resource.

This work falls into the arts, particularly in the area of landscaping or garden design. It is for a general audience. The introduction states, "The Japanese Garden Database is intended as a repository of information on the historical and contemporary gardens of Japan as well as the gardens located outside Japan that have been inspired by the culture. It is a nonprofit, educational website that seeks to provide information on a selection of outstanding examples of garden art found in Japan while juxtaposing a diversity of media related to them. This juxtaposition is intended to bring about fresh insight to a body of discourse that can often be mired in romanticized and exoticized notions of Asia and the cultures therein."

B. Identify concepts in:

1. Title and subtitle.

   The Japanese Garden Database

   Note that the title contains the words *Japanese, Garden,* and *Database.* (Does this website fit the definition of a *database?*)

2. Table of contents, chapter titles, section headings, or equivalents.

   The section headings include: private gardens, JOJG (*Journal of Japanese Gardening*) articles, web articles, features archive, plants, books, designers, suppliers, organizations, biographies, FAQ, glossary, timeline, and links.

   Some of these are helpful for aboutness; some are not.

3. Preface, introduction, and/or first chapter.

   The introduction equivalent is the AboutJGarden.org page. It frequently uses several terms such as gardens, plants, designers, and Japanese. Note that it states that the site was originally intended for professionals in the field of landscape architecture and garden history but has been broadened to be made available to the general public through the Internet.

4. Illustrations and their captions.

   Note that illustrations are of various gardens, some at varying times of year, and that there are also topographical and location maps. However, these maps do not constitute a large proportion of the work.

5. Conclusion (e.g., last chapter, section, etc.).

   None in this resource.

C. Identify names used as subject concepts.

The personal names involved are names of garden designers. There are too many to include at the summarization level. A few corporate names are given in lists, such as lists of organizations or lists of related sites; the work is not about these corporate bodies, however.

D. Identify role(s) of any geographic name(s) present.

The geographic area Japan provides the context for the topic of this work. The work is not about Japan, but the gardens discussed are of a type originally designed in Japan. Only *some* of the individual gardens discussed in this site actually are in Japan.

E. Identify chronological elements.

The time-periods involved here are from several centuries BCE to the present, much too long a time to be meaningful in a subject heading string.

F. Identify form of the resource being analyzed.

This resource consists of many different kinds of information: text, pictures, maps, and the like. Note that the word *database* is included in the subtitle, but the entire work is not a database. Much of the resource comprises internal web pages and links to outside resources. A large portion of this resource is dedicated to lists of articles about Japanese gardens (i.e., a bibliography).

II. Construct an aboutness statement.

A. Describe what the resource is about.

This resource is about the history and design of Japanese-style gardens (as a concept) and also the history and design of major Japanese-style gardens in Japan and around the world. Much of the resource contains text and images of gardens, but large portions comprise a bibliography of resources related to the topic and a database of important Japanese-style gardens.

B. Review the resource again. Look for support of your ideas. Refine the statement as needed.

C. Identify terms from the aboutness statement to be searched in the controlled vocabulary.

Japan, Japanese, Japanese-style

Gardens

History, Historic

Design

Database

Bibliography

D. Sketch out a rudimentary hierarchy (discipline/sub-discipline/concept/topic, etc.) into which the aboutness concepts fall.

Arts / design / landscaping / gardens / Japanese

III. Translate the aboutness into controlled vocabulary

A. Search the chosen terms in the controlled vocabulary. Translate the terms into specific headings from the controlled vocabulary list.

LCSH:

Gardens, Japanese—History

Gardens, Japanese—Design

Historic gardens—Japan

Gardens, Japanese—Bibliography

Gardens, Japanese—Databases

B. Using the hierarchy as a guide, search the classification for the most appropriate class(es) to describe the resource's aboutness. Convert the aboutness into specific classification notation(s).

LC Classification:	**SB458**
Dewey Decimal Classification:	**712.0952**

C. Review your choices and compare to similar items.

How did you do? Does this resource fit with others around it? Would you search for it using these words? If you searched using these words, would you be happy with this item?

APPENDIX B: ARRANGEMENT OF PHYSICAL INFORMATION RESOURCES IN LIBRARIES

Arrangement of physical information resources has been of concern to information centers for centuries. Clay tablets, papyrus scrolls, parchment, and eventually paper resources all had to be arranged. In the twentieth century, various new media (e.g., films, photographs, sound recordings) appeared, needing eventually to be arranged. Since the early 1980s, there has been concern about arrangement of electronic resources contained in physical packaging (e.g., CD-ROMs, videodiscs, music CDs).

Physical resources are often arranged by *call number*.[1] A call number is a notation on a resource that matches the same notation in the metadata record. It is the number used to "call" for an item in closed stacks —thus, the source of the name *call number*. A call number usually consists of at least two lines on a label placed on the outside of the packaging. The top line is usually a classification notation. With long notations, the classification notation may continue onto a second line. The next line is usually a *cutter number*.

Cutter numbers were devised by Charles A. Cutter more than a century ago for the purpose of creating a logical sub-arrangement of resources under classification notations. In most instances, the most logical sub-arrangement is alphabetical by the primary access point, whether it is an author, creator, or title. In essence, cuttering alphabetizes all of the works that fall under the same classification notation. Cutter devised a table in which letters of the alphabet were listed in one column and equivalent numerals were listed in an adjoining column (see Figure B.1).

Using the following example from "F" in Cutter's table,

Fonti	**684**
Fontr	**685**
Foo	**668**

a Cutter number for an author named Fontrain might be F685d. The small "d" is a *work mark* assigned to help keep the works of a particular author who writes on the same subject in alphabetical order. The work mark usually stands for the first word (that is not an article) of the title of the work; although for biographies the work is usually *cuttered* for the name of the person the work is about, and the work mark stands for the last name of the author of the biography. This kind of Cutter number is most often used with *Dewey Decimal Classification* (DDC) notations.

Goetz	611	Goun	711	Greene J	811	Guald	911
Gof	612	Goup	712	Greene S	812	Gualt	912
Gog	613	Gour	713	Greenh	813	Guan	913
Goh	614	Gourd	714	Greenl	814	Guar	914
Goi	615	Gourg	715	Greeno	815	Guari	915
Gois	616	Gouri	716	Greenw	816	Guarn	916
Gol	617	Gous	717	Gref	817	Guas	917
Gold	618	Gout	718	Greg	818	Guat	918
Goldi	619	Gouv	719	Gregg	819	Guaz	919
Goldo	621	Gov	721	Gregori	821	Gub	921
Golds	622	Gow	722	Gregory	822	Gud	922
Goldsc	623	Gower	723	Gregory M	823	Gudm	923
Goldsm	624	Goy	724	Grei	824	Gue	924
Gole	625	Goz	725	Grel	825	Guel	925
Goli	626	Gr	726	Gren	826	Guen	926
Golo	627	Grab	727	Greni	827	Gueno	927
Golov	628	Graber	728	Grenv	828	Guep	928
Golt	629	Grac	729	Grep	829	Guer	929
Gom	631	Graci	731	Gres	831	Guere	931
...		...		...		...	

Figure B.1. A Representation of a *Cutter Table*.

Library of Congress Classification (LCC) also uses cutter numbers (with a small "c") to differentiate resources that share the same classification notation (see Figure B.2). LC borrowed Charles Cutter's idea but created its own table that takes only a few lines. With LCC, a third line for a call number is often the date. (Dates are usually used in DDC call numbers only for another edition of a work for which an institution already has an edition.)

LC Cutter Table

(1)	After initial ***vowels***								
	for the second letter:	b	d	l-m	n	p	r	s-t	u-y
	use number:	2	3	4	5	6	7	8	9
(2)	After initial letter ***S***								
	for the second letter:	a	ch	e	h-i	m-p	t	u	w-z
	use number:	2	3	4	5	6	7	8	9
(3)	After initial letters ***Qu***	a	e	i	o	r	t	y	
	for the third letter:	3	4	5	6	7	8	9	
	use number:								
	For initial letters ***Qa-Qt***								
	use numbers: **2-29**								
(4)	After other initial ***consonants***								
	for the second letter:	a	e	i	o	r	u	y	
	use number:	3	4	5	6	7	8	9	
(5)	For ***expansion***								
	for the letter:	a-d	e-h	i-l	m-o	p-s	t-v	w-z	
	use number:	3	4	5	6	7	8	9	

In the examples below, Cutters without asterisks conform to the table [above]. Cutters with single asterisks have been developed when a range of letters in the table has been provided e.g., **l-m**. Cutters with two asterisks have been developed when the second letter is not explicitly stated in the table, e.g., **h** after an initial consonant. In most cases, Cutters must be adjusted to file an entry correctly and to allow room for later entries.

Vowels		**S**		**Q**		**Consonants**	
IBM	.I26	Sadron	.S23	**Qadduri	.Q23	Campbell	.C36
Idaho	.I33	**Scanlon	.S29	**Qiao	.Q27	Ceccaldi	.C43
*Ilardo	.I4	Schreiber	.S37	Quade	.Q33	**Chertok	.C48
*Import	.I48	*Shillingburg	.S53	Queiroz	.Q45	**Clark	.C58
Inman	.I56	*Singer	.S57	Quinn	.Q56	Cobblestone	.C63
Ipswich	.I67	Stinson	.S75	Quorum	.Q67	Cryer	.C79
*Ito	.I87	Suryani	.S87	Qutub	.Q88	Cuellar	.C84
*Ivy	.I94	*Symposium	.S96	**Qvortrup	.Q97	Cymbal	.C96

Figure B.2. The LC Cutter Table.

Examples of call numbers are

DDC: 378.4

F685d

LCC: QE22.D25
S65
1997

There are also alphabetical arrangements, accession order arrangements (i.e., numbers assigned to items in the order in which they arrive), and fixed location arrangements for physical information resources. In public and school libraries, in particular, fiction is often grouped together, sub-arranged in alphabetical order by author (sometimes with a prefix indicating *fiction,* such as **F** or **Fic**). Other uses for alphabetical order include arrangements of biographies and serials. Biographies may be arranged in alphabetical order by the last name of the person the biography is about (sometimes with **B** as a prefix indicating *biography*). And, as discussed in Chapter 11, serials are often arranged in alphabetical order by title. Accession order arrangements may be used for materials waiting to be cataloged. They may also be used for fixed location settings, such as remote storage.

In most institutions there is more than one sequence of the classification scheme or other arrangement used (e.g., more than one A–Z sequence, if using LCC, or more than one 000–999 sequence, if using DDC). There is often a reference collection, for example, that gathers together information resources from all parts of the classification scheme. In some libraries the prefix **R** or **Ref** is used to indicate an item's inclusion in the reference collection. These are kept separately from regular circulating collections. In academic libraries there are often collections of reserve items for use for certain courses. In many libraries some kinds of information resources are separated by format. Microforms, DVDs, sound recordings, and the like, are arranged in their own groups with their own sequences. This is often for housing and preservation purposes. Some media centers have tried to interfile all formats in classification order, but the idea has never really caught on.

NOTE

1. For more explanation of call numbers, see Chapter 19 of Daniel N. Joudrey, Arlene G. Taylor, and David P. Miller, *Introduction to Cataloging and Classification,* 11th ed. (Santa Barbara, CA: Libraries Unlimited, 2015).

APPENDIX C: ARRANGEMENT OF METADATA DISPLAYS

For most of the twentieth century, catalogs were in card, book, or microform formats. In card catalogs, the order of surrogate records was by a filing arrangement achieved by humans. Persons doing the filing had rules to follow that would lead to an arrangement that was deemed to be conducive to helping users find what they were looking for. In early book catalogs, the same was true because book catalogs often were made from cards or slips created by catalogers and placed in order by filers.

As computers entered the picture, book and then microform catalogs began to be created in electronic databases where the actual filing was done according to computer algorithms. Finally, Online Public Access Catalogs (OPACs) came into use. The records were stored sequentially or randomly within the computer, but in order to display the results of a search to a user, the subset of records retrieved had to be arranged and then displayed on the screen in some order. Early computer algorithms dictated an order of characters: sometimes numerals preceded letters, sometimes not; and there was (and still is) great difficulty in even displaying diacritical marks, let alone arranging them to display in some meaningful order.

Search results in OPACs and other databases continue to be arranged on screen by computer algorithm. A few systems have been developed to make use of MARC tags and subfield codes to create more logical arrangements. Whatever the algorithm, no computer can make up for a typographical error the way a filer can. In a card catalog, an alert filer would see a *typo* such as *Form one to zero* and file it properly as *From one to zero.* In a computer, which would see *form* as a properly spelled word in any spelling checker, the title with the typo would be arranged before all other entries for the correctly spelled title. Intervening between *form* and *from* would be titles beginning with words such as *formal*, *formation, Franciscan*, *fraud*, *frock*, and many hundreds of others. (Note: Retrieval of known items is much more difficult when a typographical error is involved.)

FILING HISTORY

Charles A. Cutter included filing rules in his cataloging code, in addition to his rules for description, name headings, and subject headings.[1] There have been separate filing codes, not connected with cataloging rules, ever since Cutter. Earlier filing codes reflected the influence of the classified catalog. They presorted catalog entries into categories. For example, entries beginning with the word *orange* were sorted into: personal names —separating single surname entries (e.g., **Orange, Carolyn**) from compound surname entries (e.g., **Orange-Keysville, James**); geographic entities (e.g., **Orange County (Calif.)**); corporate bodies (e.g., **Orange and Rockland Utilities, Inc.**); subject headings—separating the fruit (e.g., **Orange peel**) from the color (e.g., **Orange azalea**); and titles (e.g., ***Orange Bear reader***). Even though the rules were for *dictionary* catalogs, the categories were to be filed one after another instead of being interfiled alphabetically. The earlier codes also reflected the many variant local practices that existed. The 1942 *A.L.A. Rules for Filing Catalog Cards*[2] had many rules with two or three correct alternatives for a particular filing dilemma. One library could use rules 1.a. and 2.b., while another library could use rules 1.b. and 2.a. Both could claim to be following the American Library Association (ALA) filing rules. Users going from one library to another had to learn new filing rules for each catalog.

By the mid-1960s frustration with the rules had peaked. People in ALA decided to create a consistent code of filing rules derived from one basic principle. The resulting 1968 *ALA Rules for Filing Catalog Cards*[3] recommended straight alphabetical order, but there were exceptions. For example, personal surname entries (for single surnames only) were to be arranged before other entries beginning with the same word (e.g., **Love, Harold G.** was filed before ***Love and beauty***). Also, a filer had to spell out (mentally) numerals and abbreviations in the language of the item and then file the card in the place where the spelled-out form would go in the catalog (e.g., ***1984*** was filed as ***Nineteen eighty-four***).

In 1980 ALA again published a code of filing rules,[4] and this time they were so different that they were not identified as a third edition. They called for straightforward filing according to principles, but still, alphabetical order was not absolute. A brief review of these rules serves to highlight some of the problems that keep people from finding what they want in computer displays, especially when a displayed set takes up more than one or two screens.

GENERAL RULES FOR ARRANGEMENT

Traditional arrangement in catalogs is *word-by-word.* This means that everything beginning with a particular word should precede other entries beginning with a word that has the same beginning letters as the first word (e.g., *New York* files before *Newark*). Not all information retrieval tools use this arrangement. Some encyclopedias and dictionaries, for example, use *letter-by-letter* arrangement. The difference is that in word-by-word arrangement, a space between words is treated according to the principle "nothing files before something." That is, a space is "nothing," and it should be filed before a character, which is "something." According to the *ALA Filing Rules,* spaces, dashes, hyphens, diagonal slashes, and periods are all considered to be "nothing." Arrangements using the letter-by-letter approach ignore spaces and some of the punctuation marks just mentioned, and the entry files as if it is all run together into one word (e.g., *New York* is treated as *Newyork* and follows *Newark*). A longer example may assist in clarifying the distinction:

Word-by-Word	Letter-by-Letter
A book about myself	**A book about myself**
Book bytes	**Bookbinding**
Book-making (Betting)	**Book bytes**
Book of bells	**Booker T. Washington**
Book reports	**Booker, William, 1905--**
Bookbinding	**Bookfinder**
Booker T. Washington	**Bookkeeping made simple**
Booker, William, 1905--	**Book-making (Betting)**
Bookfinder	**Book of bells**
Bookkeeping made simple	**Book reports**
Books that changed the world	**Booksellers and bookselling**
Booksellers and bookselling	**Books that changed the world**

Another principle of arrangement is that numerals precede letters. Formerly, numerals were filed as spoken and spelled out, for example,

Twenty-four dramatic cases . . .
24 ways to . . .
XXIVth Congress of . . .

However, in the 1980 filing rules the difficulties of mentally spelling out *24* in French, German, Russian, and so on, for both filers and users were recognized. Numerals were to go first, but in numerical order, and Roman numerals filed with Arabic numerals. Thus, *1/2* (the fraction) was to file before *1* (the first integer). But computers have difficulty with this. The computer sees *1/2* as *one-slash-two,* not *one-half.* In the following example one can see that numerical order for a computer is based upon an arrangement of numerals by whatever is the first digit, then the second digit, and so on, rather than by the value of the number, and Roman numerals are seen as letters:

Manual Filing	Machine Arrangement
6 concerti grossi	**10 times a poem**
9 to 5	**1984**
10 times a poem	**6 concerti grossi**
XIXth century drawings	**9 to 5**

90 days to a better heart	**90 days to a better heart**
1984	**XIXth century drawings**

Among other principles of manual filing are: letters in the English alphabet precede letters of non-Roman alphabets; an ampersand (&) may be ignored, or optionally, may be spelled out in its language equivalent; and punctuation, nonalphabetic signs, and symbols are ignored. Computers generally treat these the same way except that they cannot follow the option of spelling out ampersands in their language equivalents. Ampersands are problematic for retrieval because a user who has heard a title, but has not seen it, does not know whether *and* is spelled out or represented by an ampersand and usually does not know to try the other way if one way does not work. Catalogers, however, in order to ensure access, often make additional title access points for the spelled-out form when ampersands or numerals have been used in titles.

FILING/DISPLAY DILEMMAS

There are quite a few situations that cannot be resolved by human intellect as was done in the past. It seems impossible to program computers to handle all these situations so that the outcome is logical. For example, when titles are transcribed from preferred sources of information in the process of creating metadata, the title is taken character by character as it appears. In titles that start with names beginning with prefixes, for example, the names are sometimes written with a space following the prefix (e.g., DeGaulle and De Gaulle); or, titles may start with or contain words that began as two words and are in the process of becoming one word (e.g., *on line, on-line,* and *online*). This means that some titles that begin with or contain the same words according to human understanding appear to a computer to contain different words. Here is another case where a user, having heard a title spoken, would not know which way to look for it.

Another spacing problem comes from punctuation marks. As mentioned above, the *ALA Filing Rules* says that spaces, dashes, hyphens, diagonal slashes, and periods are all considered to be "nothing." However, this is not true in all computer systems. Such marks might be replaced by a space, or they might not even be replaced by a space, resulting in words being run together inappropriately (e.g., *surrogate/metadata records* might become *surrogatemetadata records*).

Computers also have difficulty with abbreviations. A human or a natural language processing system can usually tell whether *Co.* means *County* or *Company*; whether *St.* means *Street* or *Saint*; whether *Dr.* means *Doctor* or *Drive.* Most systems cannot tell the difference, however, and arrangement is simply done by the letters that are there. The resulting retrieval challenge is that even if users know the titles they are searching for word-for-word, if they have only heard the titles and have not seen them, they do not know if certain words are abbreviated or spelled out.

Dates in subject headings present yet another challenge. In traditional manual practice, dates are to be arranged in chronological order. With *Library of Congress Subject Headings* (LCSH), however, some dates are preceded by a verbal phrase identifying a name for the particular period of time (e.g., **United States—History—Revolution, 1775–1783**). These headings can be placed in chronological order by a human, but a computer cannot see *1775–1783* until it has arranged *Revolution* preceding or following all other numerals. The following table contains a comparison of human versus machine arrangement as applied to dates:

Dates (Manual Filing)	Dates (Machine Arrangement)
United States—History—Revolution, 1775–1783	**United States—History—1800–**
United States—History—1800–	**United States—History—1801–1809**
United States—History—1801–1809	**United States—History—1900–**
United States—History—War of 1812	**United States—History—1945–**
United States—History—Civil War, 1861–1865	**United States—History—Civil War, 1861–1865**
United States—History—1900–	**United States—History—Revolution, 1775–1783**
United States—History—1945–	**United States—History—War of 1812**

Some machine systems place numerals after letters, but in any case the letters would be together and the numerals would be together. How many users know that in a long list of time periods for the history of a country, the named periods come last or the dates alone come last?

Initials and acronyms present yet another arrangement and retrieval challenge. If they are written with periods or spaces between them, they are filed as if each letter is a word (e.g., *A B C* files as if the first word is *A,* the second is *B,* and the third word is *C*; while *ABC* files as a single word). For example:

A.A.
A.A.U.W.
A apple pie
A B C programs
AAA
Aabel, Marie
Abacus calculating
ABCs of collecting

Users are at a loss to know whether periods or spaces have been used. This kind of situation requires the assistance of catalogers to add access points both with and without spaces.

The final arrangement problem to be discussed here is the one surrounding initial articles and elisions. Articles (*a, an, the,* and their equivalents in other languages) that come at the beginning of an access point are supposed to be ignored in filing. This can only happen in all languages if the system is using an encoding scheme like MARC, wherein a human can provide in an indicator the number of characters that the computer should ignore before beginning the arrangement order. Even in the MARC format there is no provision for indicating articles for every possible title access point. In MARC 21, for example, uniform titles and titles in subfield *t* of several access point fields do not have indicators for articles. In addition, if a system is programmed to give access to subtitles, those beginning with articles will be arranged under the articles.

A system cannot just have a stopword list of all articles in all languages. First, some articles are ordinary words in other languages (e.g., the German *die*). Second, if the article is part of a proper name, it should be arranged under the article. For example:

Los angeles custodios *[title in A]*
Los Angeles in fiction *[title in L]*
Los Angeles Bar Association *[corporate body in L]*

Elisions (substitution of an apostrophe for a letter or letters when running two words together, e.g., *they'll* for *they will*) present a similar problem, especially when they begin an access point. In the following example an elided article begins each access point:

L'enfant abandonee *[title in E]*
L'Enfant, Edouard *[person in L]*

The main concern in the issues discussed above is that in online systems, when searchers retrieve responses that take more than two screens to display, they have to understand the arrangement. If they think that responses are in alphabetical order, and if they expect, for example, an acronym to be at the beginning of the listing, they might not even go to the screen that actually has the entry. Although arrangement is done by the computer system, some programming can be done by humans to enable displays to be more predictable.

Arrangement of retrieved metadata in response to a search is still evolving. The first online catalogs were automated card catalogs, but they lacked the sophisticated filing arrangements that could be accomplished in card catalogs. Display of search results has had many improvements, but problems still exist. An appropriate display is highly dependent upon system design, as discussed in Chapter 4.

NOTES

1. Charles A. Cutter, *Rules for a Dictionary Catalog,* 4th ed. (Washington, DC: Government Printing Office, 1904; reprint, London: Library Association, 1962), 12.

2. *A.L.A. Rules for Filing Catalog Cards* (Chicago: American Library Association, 1942).

3. *ALA Rules for Filing Catalog Cards,* 2nd ed. (Chicago: American Library Association, 1968).

4. *ALA Filing Rules* (Chicago: American Library Association, 1980).

APPENDIX D: EAD3 ENCODED FINDING AID FOR COLLECTION DESCRIBED IN FIGURE 3.10

```
<?xml version="1.0" encoding="UTF-8"?>
<?xml-model href="ead3.rng" type="application/xml"
schematypens="http://relaxng.org/ns/structure/1.0"?>
<ead xmlns="http://ead3.archivists.org/schema/">
    <control>
        <recordid instanceurl="http://beatleyweb.simmons.edu/collectionguides
          /CharitiesCollection/ CC030.xml">CC030</recordid>
        <filedesc>
            <titlestmt>
                <titleproper>Guide to the Industrial School for Girls (Dorchester, Boston, Mass.)
                  records, 1873-1934</titleproper>
                <author>Processed by: Molly Tierney; supervised by: Jason Wood; machine-readable
                  finding aid created by: Katie Sallade</author>
            </titlestmt>
            <publicationstmt>
                <publisher>Simmons College Archives, Simmons College</publisher>
                <address>
                    <addressline>Boston, MA, U.S.A.</addressline>
                </address>
                <p>
                    <date>© 2012</date>
                    Simmons College. All Rights Reserved.
                </p>
            </publicationstmt>
            <notestmt>
                <controlnote>
                    <p>Part of the LEADS project, Simmons College Archives and the Simmons College
                      Graduate School of Library and Information Science.</p>
                </controlnote>
            </notestmt>
        </filedesc>
        <maintenancestatus value="new"/>
        <maintenanceagency>
            <agencyname>Simmons College. Archives</agencyname>
```

```
    </maintenanceagency>
    <languagedeclaration>
        <language langcode="eng"/>
        <script scriptcode="Latn"/>
    </languagedeclaration>
    <conventiondeclaration>
        <citation>DACS</citation>
    </conventiondeclaration>
    <maintenancehistory>
        <maintenanceevent>
            <eventtype value="created"/>
            <eventdatetime>2012</eventdatetime>
            <agenttype value="human"/>
            <agent>Katie Sallade</agent>
        </maintenanceevent>
    </maintenancehistory>
</control>

<archdesc level="collection" relatedencoding="MARC">
    <did>
        <head>Descriptive Summary</head>
        <unittitle encodinganalog="245">Industrial School for Girls (Dorchester, Boston, Mass.)
          records</unittitle>
        <unitdatestructured label="Dates:" unitdatetype="inclusive" encodinganalog ="245">
            <daterange>
                <fromdate>1873</fromdate>
                <todate>1934</todate>
            </daterange>
        </unitdatestructured>
        <unitid countrycode="US" repositorycode="MBSi-LS" encodinganalog="099">CC
          30</unitid>
        <origination label="Creator:">
            <corpname rules="AACR" encodinganalog="110">
                <part>Industrial School for Girls (Dorchester, Boston, Mass.)</part>
            </corpname>
        </origination>
        <langmaterial encodinganalog="546">
            <language langcode="eng">English</language>
        </langmaterial>
        <physdescstructured label="Quantity:" physdescstructuredtype="spaceoccupied"
          coverage="whole" encodinganalog="300">
            <quantity>0.5</quantity>
            <unittype>linear feet</unittype>
        </physdescstructured>
        <physdesc>1 manuscript box</physdesc>
        <repository>
            <corpname>
                <part>Simmons College (Boston, Mass.).</part>
```

```xml
            <part>Archives.</part>
        </corpname>
    </repository>
    <physloc>Collection may be stored offsite. Please contact Archives staff for more
      information.</physloc>
    <abstract encodinganalog="545">The Industrial School for Girls was established in 1853
      and incorporated 1855 in the names of Lucretia O. Everett and Maria Greenwood. It
      moved from Winchester, Massachusetts, to Dorchester, Massachusetts, in 1858, located
      at 232 Centre Street. It provided a home and training school for various branches of
      housework to develop the habits and principles to become upright, self-supported
      women. By the mid-1940s the Industrial School for Girls had evolved into the Everett
      House located at the same place, and in the 1950s the New England Home for Little
      Wanderers acquired it.</abstract>
    <abstract encodinganalog="520">The Industrial School for Girls collection contains the
      annual reports of the Board of Managers, arranged chronologically with a concentration
      of reports from 1873 to 1934. Also, there is a paper read at the 50th anniversary of the
      Dorchester Industrial School on June 7, 1904, as well as a bound copy of <title
      render="italic"><part>Suggestions to Visitors of Dependent Children</part></title>,
      1874.</abstract>
</did>

<accessrestrict encodinganalog="506">
    <head>Access Restrictions</head>
    <p>Collection is open.</p>
</accessrestrict>
<userestrict>
    <head>Copyright Notice</head>
    <p>Copyright for materials resides with the creators of the items in question, unless
      otherwise designated.</p>
</userestrict>
<prefercite>
    <head>Preferred Citation</head>
    <p>[Identification of item: description and date], Industrial School for Girls (Dorchester,
      Boston, Mass.) records, CC 30, Simmons College Archives, Boston, MA, USA.</p>
</prefercite>
<acqinfo encodinganalog="541">
    <head>Acquisitions Information</head>
    <p>Transferred from the Simmons College School of Social Work Library, 1991</p>
    <p>Accession number: 2002.178</p>
</acqinfo>
<processinfo>
    <head>Processing Information</head>
    <p>Processed by Molly Tierney, December 2002</p>
    <p>Supervised by Jason Wood</p>
    <p>This collection guide was encoded as part of the LEADS project by Katie Sallade,
      November 2012</p>
</processinfo>
<userestrict>
```

<head>Publishing Permission</head>
<p>Please contact the College Archivist with requests to publish any material from the collection.</p>
</userestrict>
<bioghist>
<head>Organizational History</head>
<p>The Industrial School for Girls was established in 1853 and incorporated 1855 in the names of Lucretia O. Everett and Maria Greenwood. It moved from Winchester, Massachusetts, to Dorchester, Massachusetts, in 1858, located at 232 Centre Street. It provided a home and training school for various branches of housework to develop the habits and principles to become upright, self-supported women. They accepted on average 25 students through an application process, . . . girls between 10 to 14 years of age, whose family or friends are unable or unfit to care for them. The original age of acceptance was between the ages of 6 to 10, which gradually rose with modern social standards. Reductions of boarding charges were made when relatives could not pay. The girls attend public schools (this started in 1881) and Congregational church. They would go out to earn their living as soon as able under the immediate care of the head of household, usually in country families, and each one, unless returned to relatives, would be supervised under the care of one of the managers. By the mid-1940s the Industrial School for Girls had evolved into the Everett House located at the same place. And in the 1950s the New England Home for Little Wanderers acquired it.</p>
<p>Information taken from <title render="italic"><part>Directory of Charitable and Beneficent Organizations</part></title>, Boston, 1907 and 1940 or the <title render="italic"><part>Report of the Board of Managers</part></title>, 1926 and the finding aid for CC 6,<title render="doublequote"><part>Guide to the New England Home for Little Wanderers records</part></title>.</p>
</bioghist>
<scopecontent>
<head>Collection Overview</head>
<p>The Industrial School for Girls records consist of the annual reports of the Board of Managers, arranged chronologically with a concentration of reports from 1873 to 1934. Also, there is a paper read at the 50th anniversary of the Dorchester Industrial School June 7, 1904, which has a summation of the first 50 years of the School and a bound copy of Suggestions to Visitors of Dependent Children, 1874 which provides guidance for guardians of the girls.</p>
<p>There has been an annual report printed every year except in 1858 when they were moving from Winchester to Dorchester. The reports cover officers' positions, brief history of the School, student population, admission application statistics, requests for student help placement, expense reports, list of subscribers for previous year, donations made and by-laws of the organization. Missing are the years: 1919, 1923, 1925, 1928, 1930, and 1932.</p>
</scopecontent>
<arrangement>
<head>Collection Arrangement</head>
<p>Collection is arranged into 3 series:</p>
<list listtype="unordered">
<item><ref target="cc030s1">Annual Reports</ref></item>
<item><ref target="cc030s2">Manual</ref></item>

```xml
        <item><ref target="cc030s3">50th Anniversary Paper</ref></item>
    </list>
</arrangement>
<controlaccess>
    <head>Online Catalog Headings</head>
    <p>These and related materials may be found under the following headings in online
      catalogs.</p>
    <list listtype="unordered">
        <item>
            <subject encodinganalog="650" source="lcsh">
                <part>Charities</part>
                <part>Massachusetts</part>
                <part>Boston</part>
            </subject>
        </item>
        <item>
            <geogname encodinganalog="651" source="lcsh">
                <part>Dorchester (Boston, Mass.)</part>
            </geogname>
        </item>
        <item>
            <corpname encodinganalog="610" source="lcnaf">
                <part>Industrial School for Girls (Dorchester, Boston, Mass.)</part>
            </corpname>
        </item>
        <item>
            <corpname encodinganalog="610" source="lcnaf">
                <part>New England Home for Little Wanderers</part>
                <part>History</part>
            </corpname>
        </item>
        <item>
            <geogname encodinganalog="651" source="lcsh">
                <part>Winchester (Mass.)</part>
            </geogname>
        </item>
    </list>
</controlaccess>
<relatedmaterial>
    <p>Part of the School of Social Work Library Charities Collection.</p>
</relatedmaterial>

<dsc dsctype="combined">
    <head>Detailed Description of the Collection</head>
    <c01 level="series" id="cc030s1">
        <did>
            <unittitle label="Series I">Annual Reports</unittitle>
            <unitdate>1873-1934 </unitdate>
```

```xml
        <physdesc>3 folders</physdesc>
    </did>
    <scopecontent>
        <p>This series contains the Industrial School for Girls' annual reports printed each
          year except in 1858. The reports cover officers' positions, brief history of the
          School, student population, admission application statistics, requests for student
          help placement, expense reports, list of subscribers for previous years, donations
          made, and by-laws of the organization. Missing years include: 1919, 1923, 1925,
          1928, 1930, and 1932.</p>
    </scopecontent>
    <c02>
        <did>
            <container localtype="box">1</container>
            <container localtype="folder">1</container>
            <unittitle>Bound 1873-1885, 1886-1900</unittitle>
        </did>
    </c02>
    <c02>
        <did>
            <container localtype="folder">2</container>
            <unittitle>1874-1920</unittitle>
        </did>
    </c02>
    <c02>
        <did>
        <container localtype="folder">3</container>
        <unittitle>1921-1934</unittitle>
        </did>
    </c02>
</c01>
<c01 level="series" id="cc030s2">
    <did>
        <unittitle label="Series II">Manual</unittitle>
        <unitdate>1879</unitdate>
        <physdesc>1 folder</physdesc>
    </did>
    <scopecontent>
        <p>This series contains a bound copy of <title render="italic"><part>Suggestions to
          Visitors of Dependent Children</part></title>, 1874 which provides guidance for
          guardians of the girls.</p>
    </scopecontent>
    <c02>
        <did>
            <container localtype="folder">4</container>
            <unittitle>1879</unittitle>
        </did>
    </c02>
</c01>
```

```xml
            <c01 level="series" id="cc030s3">
                <did>
                    <unittitle label="Series III">50th Anniversary Paper</unittitle>
                    <unitdate>1904</unitdate>
                    <physdesc>1 folder</physdesc>
                </did>
                <scopecontent>
                    <p>This series contains a paper read at the 50th anniversary of the first 50 years of
                      the Dorchester Industrial School on June 7, 1904, which has a summation of the
                      first 50 years of the School.</p>
                </scopecontent>
                <c02>
                    <did>
                        <container localtype="folder">5</container>
                        <unittitle>1904</unittitle>
                    </did>
                </c02>
            </c01>
        </dsc>
    </archdesc>
</ead>
```

APPENDIX E: BIBFRAME 2.0 RECORD

(Source Record: Library of Congress Online Catalog-Record Number 18660797.)

This is the BIBFRAME coding for the resource illustrated in Figures 6.3 and 6.4.

```
<rdf:RDF xmlns:bf = "http://id.loc.gov/ontologies/bibframe/" xmlns:bflc =
"http://id.loc.gov/ontologies/bflc/" xmlns:madsrdf = "http://www.loc.gov/mads/rdf/v1#" xmlns:rdf =
"http://www.w3.org/1999/02/22-rdf-syntax-ns#" xmlns:rdfs = "http://www.w3.org/2000/01/rdf-
schema#">
  <bf:Work rdf:about="http://bibframe.example.org/18660797#Work">
    <bf:adminMetadata>
      <bf:AdminMetadata>
        <bf:status>
          <bf:Status>
            <bf:code>c</bf:code>
          </bf:Status>
        </bf:status>
        <bflc:encodingLevel>
          <bflc:EncodingLevel>
            <bf:code>f</bf:code>
          </bflc:EncodingLevel>
        </bflc:encodingLevel>
        <bf:descriptionConventions>
          <bf:DescriptionConventions>
            <bf:code>isbd</bf:code>
          </bf:DescriptionConventions>
        </bf:descriptionConventions>
        <bf:identifiedBy>
          <bf:Local>
            <rdf:value>18660797</rdf:value>
            <bf:source>
              <bf:Source rdf:about="http://id.loc.gov/vocabulary/organizations/dlc"/>
            </bf:source>
          </bf:Local>
        </bf:identifiedBy>
        <bf:changeDate rdf:datatype="http://www.w3.org/2001/XMLSchema#dateTime">2017-06-
            07T11:12:20</bf:changeDate>
        <bf:creationDate rdf:datatype="http://www.w3.org/2001/XMLSchema#date">2015-06-
            16</bf:creationDate>
```

```
    <bf:source>
      <bf:Source>
        <rdf:type rdf:resource="http://id.loc.gov/ontologies/bibframe/Agent"/>
        <rdfs:label>DLC</rdfs:label>
      </bf:Source>
    </bf:source>
    <bf:source>
      <bf:Source>
        <rdf:type rdf:resource="http://id.loc.gov/ontologies/bibframe/Agent"/>
        <rdfs:label>DLC</rdfs:label>
      </bf:Source>
    </bf:source>
    <bf:descriptionLanguage>
      <bf:Language rdf:about="http://id.loc.gov/vocabulary/languages/eng"/>
    </bf:descriptionLanguage>
    <bf:descriptionModifier>
      <bf:Agent>
        <rdfs:label>DLC</rdfs:label>
      </bf:Agent>
    </bf:descriptionModifier>
    <bf:descriptionConventions>
      <bf:DescriptionConventions rdf:about="http://id.loc.gov/vocabulary/descriptionConventions/
          rda"/>
    </bf:descriptionConventions>
    <bf:descriptionAuthentication>
      <bf:DescriptionAuthentication rdf:about="http://id.loc.gov/vocabulary/marcauthen/pcc"/>
    </bf:descriptionAuthentication>
  </bf:AdminMetadata>
</bf:adminMetadata>
<rdf:type rdf:resource="http://id.loc.gov/ontologies/bibframe/Text"/>
<bf:language>
  <bf:Language rdf:about="http://id.loc.gov/vocabulary/languages/eng"/>
</bf:language>
  <bf:genreForm>
  <bf:GenreForm rdf:about="http://id.loc.gov/vocabulary/marcgt/bib">
    <rdfs:label>bibliography</rdfs:label>
  </bf:GenreForm>
</bf:genreForm>
<bf:classification>
  <bf:ClassificationLcc>
    <bf:source>
      <bf:Source rdf:about="http://id.loc.gov/vocabulary/organizations/dlc"/>
    </bf:source>
    <bf:classificationPortion>Z693</bf:classificationPortion>
    <bf:itemPortion>.W94 2015</bf:itemPortion>
  </bf:ClassificationLcc>
</bf:classification>
<bf:classification>
```

```
  <bf:ClassificationDdc>
    <bf:classificationPortion>025.3</bf:classificationPortion>
    <bf:edition>23</bf:edition>
    <bf:edition>full</bf:edition>
    <bf:source>
      <bf:Source rdf:about="http://id.loc.gov/vocabulary/organizations/dlc"/>
    </bf:source>
  </bf:ClassificationDdc>
</bf:classification>
<bf:classification>
  <bf:Classification>
    <bf:classificationPortion>LAN025030</bf:classificationPortion>
    <bf:source>
      <bf:Source>
        <rdfs:label>bisacsh</rdfs:label>
      </bf:Source>
    </bf:source>
  </bf:Classification>
</bf:classification>
<bf:classification>
  <bf:Classification>
    <bf:classificationPortion>LAN020000</bf:classificationPortion>
    <bf:source>
      <bf:Source>
        <rdfs:label>bisacsh</rdfs:label>
      </bf:Source>
    </bf:source>
  </bf:Classification>
</bf:classification>
<bf:contribution>
  <bf:Contribution>
    <rdf:type rdf:resource="http://id.loc.gov/ontologies/bflc/PrimaryContribution"/>
    <bf:agent>
      <bf:Agent rdf:about="http://id.loc.gov/authorities/names/no2006073134">
        <rdf:type rdf:resource="http://id.loc.gov/ontologies/bibframe/Person"/>
        <bflc:name00MatchKey>Joudrey, Daniel N.,</bflc:name00MatchKey>
        <bflc:primaryContributorName00MatchKey>Joudrey, Daniel N.,
        </bflc:primaryContributorName00MatchKey>
        <bflc:name00MarcKey>1001 $aJoudrey, Daniel N.,$eauthor.</bflc:name00MarcKey>
        <rdfs:label>Joudrey, Daniel N.,</rdfs:label>
      </bf:Agent>
    </bf:agent>
    <bf:role>
      <bf:Role>
        <rdfs:label>author.</rdfs:label>
      </bf:Role>
    </bf:role>
  </bf:Contribution>
```

```
</bf:contribution>
<rdfs:label>Introduction to cataloging and classification /</rdfs:label>
<bf:title>
  <bf:Title>
    <rdfs:label>Introduction to cataloging and classification /</rdfs:label>
    <bflc:titleSortKey>Introduction to cataloging and classification /</bflc:titleSortKey>
    <bf:mainTitle>Introduction to cataloging and classification</bf:mainTitle>
  </bf:Title>
</bf:title>
<bf:content>
  <bf:Content rdf:about="http://id.loc.gov/vocabulary/contentType/txt">
    <rdfs:label>text</rdfs:label>
    <bf:source>
      <bf:Source>
        <rdfs:label>rdacontent</rdfs:label>
      </bf:Source>
    </bf:source>
  </bf:Content>
</bf:content>
<bf:subject>
  <bf:Topic rdf:about="http://bibframe.example.org/18660797#Topic650-28">
    <rdf:type rdf:resource="http://www.loc.gov/mads/rdf/v1#Topic"/>
    <rdfs:label>Descriptive cataloging.</rdfs:label>
    <madsrdf:authoritativeLabel>Descriptive cataloging.</madsrdf:authoritativeLabel>
    <madsrdf:isMemberofMADSScheme rdf:resource="http://id.loc.gov/authorities/subjects"/>
    <bf:source>
      <bf:Source>
        <bf:code>lcsh</bf:code>
      </bf:Source>
    </bf:source>
  </bf:Topic>
</bf:subject>
<bf:subject>
  <bf:Topic rdf:about="http://bibframe.example.org/18660797#Topic650-29">
    <rdf:type rdf:resource="http://www.loc.gov/mads/rdf/v1#Topic"/>
    <rdfs:label>Subject cataloging.</rdfs:label>
    <madsrdf:authoritativeLabel>Subject cataloging.</madsrdf:authoritativeLabel>
    <madsrdf:isMemberofMADSScheme rdf:resource="http://id.loc.gov/authorities/subjects"/>
    <bf:source>
      <bf:Source>
        <bf:code>lcsh</bf:code>
      </bf:Source>
    </bf:source>
  </bf:Topic>
</bf:subject>
<bf:subject>
  <bf:Topic rdf:about="http://bibframe.example.org/18660797#Topic650-30">
    <rdf:type rdf:resource="http://www.loc.gov/mads/rdf/v1#ComplexSubject"/>
```

```
      <rdfs:label>Classification--Books.</rdfs:label>
      <madsrdf:authoritativeLabel>Classification--Books.</madsrdf:authoritativeLabel>
      <madsrdf:isMemberofMADSScheme rdf:resource="http://id.loc.gov/authorities/subjects"/>
      <madsrdf:componentList rdf:parseType="Collection">
        <madsrdf:Topic>
          <madsrdf:authoritativeLabel>Classification</madsrdf:authoritativeLabel>
        </madsrdf:Topic>
        <madsrdf:Topic>
          <madsrdf:authoritativeLabel>Books</madsrdf:authoritativeLabel>
        </madsrdf:Topic>
      </madsrdf:componentList>
      <bf:source>
        <bf:Source>
          <bf:code>lcsh</bf:code>
        </bf:Source>
      </bf:source>
    </bf:Topic>
  </bf:subject>
  <bf:subject>
    <bf:Work rdf:about="http://bibframe.example.org/18660797#Work630-31">
      <rdf:type rdf:resource="http://www.loc.gov/mads/rdf/v1#Title"/>
      <madsrdf:authoritativeLabel>Resource description & access.</madsrdf:authoritativeLabel>
      <madsrdf:isMemberofMADSScheme rdf:resource="http://id.loc.gov/authorities/subjects"/>
      <bf:source>
        <bf:Source>
          <bf:code>lcsh</bf:code>
        </bf:Source>
      </bf:source>
      <rdfs:label>Resource description & access.</rdfs:label>
      <bf:title>
        <bf:Title>
          <bflc:title30MatchKey>Resource description & access.</bflc:title30MatchKey>
          <bflc:title30MarcKey>63000$aResource description & access.</bflc:title30MarcKey>
          <rdfs:label>Resource description & access.</rdfs:label>
          <bflc:titleSortKey>Resource description & access.</bflc:titleSortKey>
          <bf:mainTitle>Resource description & access</bf:mainTitle>
        </bf:Title>
      </bf:title>
    </bf:Work>
  </bf:subject>
  <bf:subject>
    <bf:Topic rdf:about="http://bibframe.example.org/18660797#Topic650-32">
      <rdf:type rdf:resource="http://www.loc.gov/mads/rdf/v1#Topic"/>
      <rdfs:label>LANGUAGE ARTS & DISCIPLINES / Library & Information Science / Cataloging &
      Classification.</rdfs:label>
      <madsrdf:authoritativeLabel>LANGUAGE ARTS & DISCIPLINES / Library & Information Science /
      Cataloging & Classification.</madsrdf:authoritativeLabel>
      <bf:source>
```

```
        <bf:Source>
          <bf:code>bisacsh</bf:code>
        </bf:Source>
      </bf:source>
    </bf:Topic>
  </bf:subject>
  <bf:subject>
    <bf:Topic rdf:about="http://bibframe.example.org/18660797#Topic650-33">
      <rdf:type rdf:resource="http://www.loc.gov/mads/rdf/v1#Topic"/>
      <rdfs:label>LANGUAGE ARTS & DISCIPLINES / Study & Teaching.</rdfs:label>
      <madsrdf:authoritativeLabel>LANGUAGE ARTS & DISCIPLINES / Study &
       Teaching.</madsrdf:authoritativeLabel>
      <bf:source>
        <bf:Source>
          <bf:code>bisacsh</bf:code>
        </bf:Source>
      </bf:source>
    </bf:Topic>
  </bf:subject>
  <bf:contribution>
    <bf:Contribution>
      <bf:agent>
        <bf:Agent rdf:about="http://id.loc.gov/authorities/names/n80050006">
          <rdf:type rdf:resource="http://id.loc.gov/ontologies/bibframe/Person"/>
          <bflc:name00MatchKey>Taylor, Arlene G., 1941-</bflc:name00MatchKey>
          <bflc:name00MarcKey>7001 $aTaylor, Arlene G.,$d1941-$eauthor.</bflc:name00MarcKey>
          <rdfs:label>Taylor, Arlene G., 1941-</rdfs:label>
        </bf:Agent>
      </bf:agent>
      <bf:role>
        <bf:Role>
          <rdfs:label>author.</rdfs:label>
        </bf:Role>
      </bf:role>
    </bf:Contribution>
  </bf:contribution>
  <bf:contribution>
    <bf:Contribution>
      <bf:agent>
        <bf:Agent rdf:about="http://id.loc.gov/authorities/names/n00031008">
          <rdf:type rdf:resource="http://id.loc.gov/ontologies/bibframe/Person"/>
          <bflc:name00MatchKey>Miller, David P. (David Peter), 1955-</bflc:name00MatchKey>
          <bflc:name00MarcKey>7001 $aMiller, David P.$q(David Peter),$d1955-
          $eauthor.</bflc:name00MarcKey>
          <rdfs:label>Miller, David P. (David Peter), 1955-</rdfs:label>
        </bf:Agent>
      </bf:agent>
      <bf:role>
```

```
      <bf:Role>
        <rdfs:label>author.</rdfs:label>
      </bf:Role>
    </bf:role>
  </bf:Contribution>
</bf:contribution>
<bf:relatedTo>
  <bf:Work rdf:about="http://bibframe.example.org/18660797#Work700-36">
    <bf:contribution>
      <bf:Contribution>
        <bf:agent>
          <bf:Agent rdf:about="http://id.loc.gov/authorities/names/n80050006">
            <rdf:type rdf:resource="http://id.loc.gov/ontologies/bibframe/Person"/>
            <bflc:name00MatchKey>Taylor, Arlene G., 1941-</bflc:name00MatchKey>
            <bflc:name00MarcKey>7001 $aTaylor, Arlene G.,$d1941-$tIntroduction to cataloging
            and classification.$iBased on (work):</bflc:name00MarcKey>
            <rdfs:label>Taylor, Arlene G., 1941-</rdfs:label>
          </bf:Agent>
        </bf:agent>
        <bf:role>
          <bf:Role rdf:about="http://id.loc.gov/vocabulary/relators/ctb"/>
        </bf:role>
      </bf:Contribution>
    </bf:contribution>
    <rdfs:label>Introduction to cataloging and classification.</rdfs:label>
    <bf:title>
      <bf:Title>
        <bflc:title00MatchKey>Introduction to cataloging and classification.</bflc:title00MatchKey>
        <bflc:title00MarcKey>7001 $aTaylor, Arlene G.,$d1941-$tIntroduction to cataloging and
        classification.$iBased on (work):</bflc:title00MarcKey>
        <rdfs:label>Introduction to cataloging and classification.</rdfs:label>
        <bflc:titleSortKey>Introduction to cataloging and classification.</bflc:titleSortKey>
        <bf:mainTitle>Introduction to cataloging and classification</bf:mainTitle>
      </bf:Title>
    </bf:title>
  </bf:Work>
</bf:relatedTo>
<bflc:relationship>
  <bflc:Relationship>
    <bflc:relation>
      <rdfs:Resource>
        <rdfs:label>Based on (work)</rdfs:label>
      </rdfs:Resource>
    </bflc:relation>
    <bf:relatedTo rdf:resource="http://bibframe.example.org/18660797#Work700-36"/>
  </bflc:Relationship>
</bflc:relationship>
<bf:hasInstance rdf:resource="http://bibframe.example.org/18660797#Instance"/>
```

```
</bf:Work>
<bf:Instance rdf:about="http://bibframe.example.org/18660797#Instance">
  <bf:issuance>
    <bf:Issuance rdf:about="http://id.loc.gov/vocabulary/issuance/mono"/>
  </bf:issuance>
  <bf:provisionActivity>
    <bf:ProvisionActivity>
      <rdf:type rdf:resource="http://id.loc.gov/ontologies/bibframe/Publication"/>
      <bf:date rdf:datatype="http://id.loc.gov/datatypes/edtf">2015</bf:date>
      <bf:place>
        <bf:Place rdf:about="http://id.loc.gov/vocabulary/countries/cau"/>
      </bf:place>
    </bf:ProvisionActivity>
  </bf:provisionActivity>
  <bf:supplementaryContent>
    <bf:SupplementaryContent>
      <rdfs:label>Index present</rdfs:label>
    </bf:SupplementaryContent>
  </bf:supplementaryContent>
  <bf:identifiedBy>
    <bf:Lccn>
      <rdf:value>2015012911</rdf:value>
    </bf:Lccn>
  </bf:identifiedBy>
  <bf:identifiedBy>
    <bf:Isbn>
      <rdf:value>9781598848571 (hardback : acid-free paper)</rdf:value>
    </bf:Isbn>
  </bf:identifiedBy>
  <bf:identifiedBy>
    <bf:Isbn>
      <rdf:value>9781598848564 (pbk : acid-free paper)</rdf:value>
    </bf:Isbn>
  </bf:identifiedBy>
  <bf:identifiedBy>
    <bf:Isbn>
      <rdf:value>9781440837456 (ebook)</rdf:value>
      <bf:status>
        <bf:Status>
          <rdfs:label>invalid</rdfs:label>
        </bf:Status>
      </bf:status>
    </bf:Isbn>
  </bf:identifiedBy>
  <rdfs:label>Introduction to cataloging and classification /</rdfs:label>
  <bf:title>
    <bf:Title>
      <rdfs:label>Introduction to cataloging and classification /</rdfs:label>
```

```
      <bflc:titleSortKey>Introduction to cataloging and classification /</bflc:titleSortKey>
      <bf:mainTitle>Introduction to cataloging and classification</bf:mainTitle>
    </bf:Title>
  </bf:title>
  <bf:responsibilityStatement>Daniel N. Joudrey, Arlene G. Taylor, and David P.
  Miller</bf:responsibilityStatement>
  <bf:editionStatement>Eleventh edition</bf:editionStatement>
  <bf:provisionActivity>
    <bf:ProvisionActivity>
      <rdf:type rdf:resource="http://id.loc.gov/ontologies/bibframe/Publication"/>
      <bf:place>
        <bf:Place>
          <rdfs:label>Santa Barbara, California</rdfs:label>
        </bf:Place>
      </bf:place>
      <bf:agent>
        <bf:Agent>
          <rdfs:label>Libraries Unlimited</rdfs:label>
        </bf:Agent>
      </bf:agent>
      <bf:date>2015</bf:date>
    </bf:ProvisionActivity>
  </bf:provisionActivity>
  <bf:provisionActivityStatement>Santa Barbara, California : Libraries Unlimited,
  [2015]</bf:provisionActivityStatement>
  <bf:extent>
    <bf:Extent>
      <rdfs:label>xxv, 1048 pages</rdfs:label>
    </bf:Extent>
  </bf:extent>
  <bf:dimensions>27 cm.</bf:dimensions>
  <bf:media>
    <bf:Media rdf:about="http://id.loc.gov/vocabulary/mediaType/n">
      <rdfs:label>unmediated</rdfs:label>
      <bf:source>
        <bf:Source>
          <rdfs:label>rdamedia</rdfs:label>
        </bf:Source>
      </bf:source>
    </bf:Media>
  </bf:media>
  <bf:carrier>
    <bf:Carrier rdf:about="http://id.loc.gov/vocabulary/carriers/nc">
      <rdfs:label>volume</rdfs:label>
      <bf:source>
        <bf:Source>
          <rdfs:label>rdacarrier</rdfs:label>
        </bf:Source>
```

```
        </bf:source>
      </bf:Carrier>
    </bf:carrier>
    <bf:seriesStatement>Library and information science text series</bf:seriesStatement>
    <bf:note>
      <bf:Note>
        <rdfs:label>Significant expansion of: Introduction to cataloging and classification. Tenth edition /
        Arlene G. Taylor ; with the assistance of David P. Miller.</rdfs:label>
      </bf:Note>
    </bf:note>
    <bf:note>
      <bf:Note>
        <rdfs:label>Includes bibliographical references (pages 1001-1022) and index.</rdfs:label>
        <bf:noteType>bibliography</bf:noteType>
      </bf:Note>
    </bf:note>
    <bf:instanceOf rdf:resource="http://bibframe.example.org/18660797#Work"/>
    <bf:hasItem>
      <bf:Item rdf:about="http://bibframe.example.org/18660797#Item050-14">
        <bf:shelfMark>
          <bf:ShelfMarkLcc>
            <rdfs:label>Z693.W94 2015</rdfs:label>
            <bf:source>
              <bf:Source rdf:about="http://id.loc.gov/vocabulary/organizations/dlc"/>
            </bf:source>
          </bf:ShelfMarkLcc>
        </bf:shelfMark>
        <bf:itemOf rdf:resource="http://bibframe.example.org/18660797#Instance"/>
      </bf:Item>
    </bf:hasItem>
  </bf:Instance>
</rdf:RDF>
```

GLOSSARY

Defined in this glossary are selected basic terms for students of information organization, including a number of terms and identifiers used in descriptive cataloging, classification, subject heading work, document indexing, archival description, and other topics treated in this text. Readers may wish to consult other sources for other terms used in the library and information professions. Some concepts are marked *obsolete* to indicate that the terms are no longer in common usage; these terms remain in the glossary, however, because they may be encountered in older LIS literature.

AACR ***(Anglo-American Cataloguing Rules).*** A set of cataloging rules, first published in 1967, for producing the descriptive metadata and name-and-title access points in a surrogate record for a resource; later editions were published in 1978, 1988, 1998, and 2002; the editions published in 1978 and later were referred to as AACR2; the creation of these rules was the result of collaboration among representatives from Australia, Canada, Great Britain, and the United States. It was replaced in 2010 by *RDA: Resource Description & Access*.

AAP. *See* **Authorized access point (AAP).**

AAT ***(Art & Architecture Thesaurus).*** A thesaurus that covers the disciplines of art and architecture.

Abbreviation. A shortened form of a word. *See also* **Acronym; Initialism.**

Aboutness. The subject of a work contained in a resource, which is translated into controlled subject languages (e.g., classification schemes, subject heading lists); includes topical aspects and also genre and form. *See also* **Conceptual analysis; Is-ness; Of-ness; Subject analysis; Translation.**

Aboutness statement. A single sentence or a short paragraph describing and summarizing one's understanding of a resource's aboutness, the genre/form, and the relationships among the important subject concepts. The aboutness statement is used to identify the terms or concepts to be searched in the controlled vocabulary. The statement can also help the cataloger to construct a rudimentary hierarchy to get a better understanding of how the concepts fit together and how they might fit into a scheme of classification.

Abstract. A condensed narrative description (i.e., summary or synopsis) of a resource that may serve as a surrogate for the resource in a retrieval system; typically created for journal articles, conference papers, individual chapters, and the like; usually created by an indexer at the time of database indexing or may be written by the author of the resource.

Abstracting. The process of creating an abstract. *See also* **Indexing.**

Access. That portion of cataloging in which access points are selected and formulated by a cataloger. *See also* **Description; Descriptive cataloging.**

Access point. Any word or phrase used to obtain information from a retrieval tool or other organized system; in cataloging and indexing, *access points* are specific names, titles, and subjects chosen by the cataloger or indexer, when creating metadata, to allow for the retrieval of the resource description. *See also* **Added entry; Authorized access point (AAP); Controlled access point; Main entry (access point); Variant access point (VAP).**

Accession log. *See* **Register.**

Accession number. A notation assigned to an information resource that is unique to the resource; the notations are often based upon the order in which resources are acquired.

Accession record. In archives and museums, a record that contains basic information about the acquisition of a collection or object. It may include an identification number, information about the donor, any associations, provenance, any information needed for insurance purposes, and so forth.

Accompanying materials. Dependent materials, such as answer books, teacher's manuals, atlases, portfolios of plates, slides, sound recordings, or computer disks, that accompany and are cataloged with the main component of the resource.

Acquisitions. The library technical services unit that, among other things, is responsible for managing orders and budgets for the collection.

Acronym. An abbreviation made up of initial letters of other words, but pronounced as its own word (e.g., *radar*). *See* **Abbreviation; Initialism.**

Added entry. *Obsolete.* In AACR2, any access point in a resource description other than the main entry; replaced in RDA by *access point*, and access points are not considered to be primary or secondary unless a work identifier is being selected. *See also* **Main entry (access point).**

Administrative metadata. Metadata created for the purposes of management, decision making, and record keeping; provides information about the technical, preservation, and storage requirements of digital resources; used for monitoring, accessing, reproducing, digitizing, and backing up digital resources.

A-G Canada. *See* **Auto-Graphics.**

Agent. A term generically referring to various roles associated with the creation and production of resources; an entity responsible for the existence, expansion, dissemination, etc., of resources (e.g., creators, contributors, and publishers of information resources). *See also* **Collective agent.**

Alphabetical catalog. A catalog with entries arranged or displayed in alphabetical order rather than according to the symbolic notation of a classification. *See also* **Classified catalog; Dictionary catalog; Divided catalog.**

Alphabetico-classed catalog. Catalog in which subject categories are used for the arrangement of resource descriptions; broad categories are subdivided by narrower categories that are placed alphabetically within each broad category.

Alternative (RDA). An instruction in RDA that offers a different approach to what was specified in the preceding instruction.

Alternative title. A second title for a resource that is joined to the first title with the word *or* (e.g., *Frankenstein, or, The Modern Prometheus*). Both titles together are considered to constitute the *title proper* of the work. *See also* **Title proper.**

Analytical description. A description of a part or parts of a larger resource (as opposed to a description of the whole resource; e.g., creating a description for one short story in a compilation). *See also* **Comprehensive description; Hierarchical description.**

Analytico-synthetic classification. *See* **Faceted classification.**

Annotation. A brief note indicating the subject matter or commenting on the usefulness of information in a particular resource.

ANSI (American National Standards Institute). A corporate body that takes responsibility for establishing voluntary industry standards; works closely with NISO.

Antonym. A term that has the opposite meaning of another term (e.g., *hot* and *cold* are antonyms). *See also* **Synonym**.

Application profile. Document that describes a community's recommended best practices for metadata creation; a formal way to declare which elements from which namespaces are used in a particular application or project or by a particular community.

APPM ***(Archives, Personal Papers, and Manuscripts).*** A standard based on AACR2 for the description of archival materials. Now replaced by DACS. See also ***Describing Archives: A Content Standard*** (DACS).

Approval plan. A method in which a library contracts with one or more vendors to receive new resources according to pre-selected profiles outlining the collection's needs.

Archival description. The process of establishing intellectual control over the holdings of an archives through the preparation of finding aids.

Archival series. A logical group of materials in an archival collection.

Archives. (1) An organization that preserves records of enduring value that document activities of organizations or persons and are accumulated in the course of daily activities; (2) materials created by organizations or persons that document the course of daily activities.

Arrangement. The placing of entities in a certain order (e.g., alphabetical, by classification, by size). In card catalogs, this activity was called *filing*.

Art & Architecture Thesaurus. *See* **AAT.**

ASCII (American Standard Code for Information Interchange). A standard code that assigns specific bit patterns to letters, numbers, and symbols; used for exchange of textual data in instances where programs are incompatible, because ASCII can be read almost universally by any computer.

Associative relationship. A relationship between terms in a controlled vocabulary that might be of interest but that is not a hierarchical or an equivalence-based relationship; often expressed as a *related term* relationship. *See also* **Related term (RT).**

Asynchronous responsibility. The situation in which the agents involved in the creation of a resource have made different kinds of contributions (e.g., author and illustrator; performer, conductor, choreographer, and producer); also called *mixed responsibility*. *See also* **Synchronous responsibility.**

Attribute. A characteristic or property of an entity, such as extent, edition statement, form, and so on, that is useful in description.

Author. A specific type of creator primarily associated with writing books or other forms of text; a person who is responsible for all or some of the intellectual content of a text. *See also* **Creator.**

Author entry. *Obsolete.* The place in a retrieval tool where a surrogate record beginning with the name of the author (i.e., creator) of an information resource may be found.

Author/Title access point. *See* **Name/Title access point.**

Authority control. The result of the process of maintaining consistency in the verbal form used to represent an access point and the further process of showing the relationships among names, works, and subjects—all for the purpose of collocation; also, the result of the process of doing authority work with or without the necessity of choosing one form of name or title or one subject term to be the authorized selection. *See also* **Authority work.**

Authority file. A grouping of authority records.

Authority record. Documentation of the decisions made during the course of authority work. An authority record contains a variety of data about the entity (e.g., person, family, corporate body, place, work, subject) being described, including all the forms used for a particular name, title, or subject and all the relationships among variants that have been identified in the process of authority work; typically the record designates one of the forms as the authorized or default one to use in catalog records.

Authority work. The process of determining and maintaining the form of a name, title, or subject concept to be used in creating access points. In the name and title areas, the process includes identifying all variant names or titles and relating the variants to the name or title forms chosen to be access points. In some cases it may also include relating names or titles to each other. In the verbal subject area, the process includes identifying and maintaining relationships among terms—relationships such as synonyms, broader terms, narrower terms, and related terms. *See also* **Authority control.**

Authorized access point (AAP). The standardized character string established in an authority record that is used to represent an entity (e.g., person, family, corporate body, place, work, subject) consistently in bibliographic descriptions. *See also* **Controlled access point; Variant access point (VAP).**

Authorized term. *See* **Preferred term.**

Auto-Graphics. A bibliographic network offering its database and services to a variety of Canadian libraries and also to a few libraries in the northeastern United States; formerly Utlas International, which became ISM/LIS, and then A-G Canada.

Axiom. In ontologies, a statement or proposition that is accepted as true.

A–Z index. *See* **Back-of-the-book index.**

Back-of-the-book index. An alphabetical list of entries for the major subjects, authors, and works referred to in an information resource. Each entry is accompanied by references or pointers (e.g., page numbers) to the locations in the resource that contain information about that entry. On the web, similar indexes may be referred to as *A–Z indexes,* with direct links to the web pages that contain information about the entries.

BIBFRAME. Bibliographic Framework Initiative, a project started by the Library of Congress to provide a replacement for MARC as the primary encoding standard for library-generated metadata. Based on linked data principles, BIBFRAME replaces self-contained bibliographic and authority records with authoritative statements about resources—statements that are encoded for use in the Semantic Web.

Bibliographic classification. *See* **Classification.**

Bibliographic control. *See* **Information organization.**

Bibliographic data. Information gathered in the process of creating resource descriptions; also refers to any discrete element in a bibliographic record. *See also* **Metadata**.

Bibliographic database. A collection of resource descriptions held in the format of a database. *See also* **Database**.

Bibliographic description. *See* **Resource description**.

Bibliographic family. A set of related resources that are derived from a common progenitor.

Bibliographic file. A grouping of bibliographic records. In a catalog or bibliographic database, a bibliographic file is distinct from, but might be linked to, one or more authority files and/or holdings files. *See also* **Authority file.**

Bibliographic Framework Initiative. *See* **BIBFRAME**.

Bibliographic identity. The concept that creators of works may use separate personae when creating different types of works. For example, Charles L. Dodgson and Lewis Carroll are two bibliographic identities used by a single person; one wrote about mathematics and the other wrote literature.

Bibliographic network. A corporate entity that has as its main resource a bibliographic database; access to the database is available for a price, and members of the network can contribute new records and download existing ones.

Bibliographic record. Catalog data in card, microform, machine-readable, or other form carrying full cataloging information for a resource. *See also* **Resource description; Surrogate record.**

Bibliographic tool. *See* **Bibliographic database; Bibliography; Catalog; Finding aid; Index; Register; Retrieval tool.**

Bibliographic universe. A concept that encompasses all instances of recorded information.

Bibliographic utility. *See* **Bibliographic network.**

Bibliography. A list of resources on a given subject, by a given author, from a particular time period or place, and the like.

Book catalog. Catalog in which resource descriptions are printed on pages that are bound into book form.

Book number. *See* **Cutter number.**

Boolean operators. The terms AND, OR, and NOT as used to construct searches in a retrieval tool that uses post-coordinate indexing and/or keyword searching.

Boolean searching. The process of searching with individual index terms or keywords that are linked with Boolean operators (either using actual operators or symbols for them, or using a system where operators are implied if not specified).

Broad classification. A method of applying a classification scheme that omits detailed subdivision of its main classes or that facilitates the use in smaller libraries of only its main classes and subdivisions. *See also* **Close classification.**

Broader term (BT). A term one level up from the term being examined in a listing where terms for subject concepts have been organized into relationships that are hierarchical. *See also* **Narrower term; Related term.**

Browsing. A process of looking, usually based on subject, at all the resources in a particular area of the stacks in order to find the items that best suit the needs of the person who is browsing. In online systems, browsing is the process of looking at all the metadata that is displayed under a certain subject, in a particular classification, or under a certain name or title.

Call number. A notation on a resource that matches the same notation in the metadata description and is used to identify and locate the item; it often consists of a classification notation and a cutter number, and it may also include a workmark and/or a date; it is the number used to "call" for an item in a closed stack library—thus the source of the name *call number. See also* **Cutter number; Work mark**.

Card catalog. Catalog in which every resource description is written, typed, or printed on cards (usually measuring 3 × 5 inches) that are placed in file drawers in a particular order (usually alphabetical or classified order).

Carrier. The physical format in which the content of a resource is stored.

Carrier type. A category reflecting the physical format of a resource (e.g., *volume*, *audio disc*, *filmstrip cartridge*).

Catalog. A type of retrieval tool; an organized compilation of bibliographic metadata or an organized set of surrogate records that represent the holdings of a particular collection and/or resources to which access may be gained. It may be arranged alphabetically, by classification notation, by subject, or, in the case of an online catalog, the display may be arranged by date or any one of several other elements.

Catalog gateway. *See* **Gateway.**

Catalog record. *See* **Bibliographic record.**

Cataloger. Person in a library or another information or cultural heritage organization who creates metadata for the resources collected by the organization and who works to maintain the system through which that metadata is made available to users; the person may also be an independent contractor. *See also* **Indexer.**

Cataloging. The creation of metadata for information resources by describing a resource, choosing appropriate access points, conducting subject analysis, assigning subject headings and classification numbers, and maintaining the system through which the cataloging data is made available. *See also* **Copy cataloging; Descriptive cataloging; Indexing; Original cataloging; Subject cataloging.**

Cataloging code. A set of rules (i.e., guidelines or instructions) for cataloging.

***Cataloging Cultural Objects* (CCO).** A content standard for communities that describe works of art, architecture, cultural artifacts, and images.

***Categories for the Description of Works of Art* (CDWA).** A set of guidelines that contains a metadata element set that is used for describing and accessing information about works of art, architecture, other material culture, and related images.

Categorization. The cognitive function that involves grouping together like entities, concepts, objects, resources, and so on.

Category. A cognitive label applied to a group of like entities.

Character-by-character filing. *See* **Letter-by-letter filing.**

Checksum. A value computed for a block of data, based on the contents of the block; used to detect corruption of the data.

Chief source of information. *Obsolete.* In AACR2, the location from which much of the information is to be taken that is used to create the descriptive part of a resource description (e.g., title page of a book, title screen of a motion picture, label on a sound recording tape or disc).

CIP (Cataloging-in-Publication). A program in which cataloging is provided by an authorized agency to the publisher or producer of a resource so that the cataloging can be issued with the resource; often the phrase is applied in the case of cataloging provided by the Library of Congress to the publishers of books.

Citation. An entry in a bibliography containing a brief description of a resource; typically contains no more than creator, title, version information, publication information, volume, issue, and page numbers.

Class. (1) In bibliographic classification, the first order of structure in a hierarchical classification, at which level major disciplines are represented. A class may incorporate one or more divisions, which in turn may incorporate one or more sections or subdivisions. (2) In ontologies, a group or category of entities with something in common.

Classical theory of categorization. Theory based on Aristotle's hypothesis that categories contain entities or concepts based on what they have in common and that categories are like containers with things either in or out of the container.

Classification. The placing of subjects into categories; in organization of information, classification is the process of determining where an information resource fits into a given hierarchy and often then assigning the notation associated with the appropriate level of the hierarchy to the information resource and to its metadata. *See also* **Enumerative classification; Faceted classification; Hierarchical classification.**

Classification notation. A set of numbers, letters, symbols, or a combination of these that is assigned to a certain concept in a classification scheme.

Classification schedule. The list of concepts found in a classification system in classified order, usually including a notation to be used for each concept.

Classification scheme. An organized framework for the systematic organization of knowledge, usually organized by subject. *See also* **Taxonomy.**

Classification table. Supplementary part of a classification scheme in which notations are assigned for concepts that can be applied in conjunction with many different topical subjects. Tables commonly exist for geographic locations, time periods, standard subdivisions (e.g., dictionaries, theory, serial publications, historical treatment), ethnic and national groups, languages, and so on.

Classified catalog. A catalog whose main part is arranged or displayed in the order of symbols, numbers, or other notations that represent the various subjects or aspects of subjects covered by the resources housed in the institution; supplementary part(s) usually exist for alphabetical indexes for creators, subject terms, and so on. *See also* **Alphabetical catalog; Dictionary catalog; Divided catalog.**

Close classification. The use of minute subdivisions to arrange materials by highly specific topics. *See also* **Broad classification.**

Closed stacks. The name given to the situation where collections are not open to public access or are limited to only a small group of users; also, resource storage areas that are accessible only to the staff of the library, archives, or other institution. *See also* **Open stacks.**

Clustering. An algorithm-based sorting technique used by retrieval tools to group similar search results together based on content (and a variety of other factors), which can then be used to display the results.

Code (as a noun). (1) A set of rules. (2) A specific designation in an encoding standard that defines and limits the kinds of data that can be stored at that point (e.g., 245 for a title statement; 650 for a subject).

Code (as a verb). The process of assigning the appropriate specified designations of an encoding standard.

Codification. The process of creating sets of rules to govern such things as the making of resource descriptions.

Coextensive subject entry. A principle of subject analysis, by which a subject heading or a set of headings covers all, but no more than, the concepts or topics covered in the resource.

Collaborative work. A single work created by two or more persons, families, or corporate bodies, in which (1) responsibility is shared and (2) the individual contributions of the creators cannot be distinguished.

Collation. *Obsolete.* In AACR2 and earlier, a statement of details about the physical description (e.g., pagination, illustrations, and size) of a book. The concept as applied to other types of resources is called *physical description* or *carrier description.*

Collection-level description. An approach to creating metadata in which two or more individual resources are described together as a unit.

Collective agent. A group of persons, acting together as a unit with a single name, that is responsible for the existence, expansion, dissemination, etc., of resources; families and corporate bodies are examples of collective agents. *See also* **Agent.**

Collective title. An inclusive title that represents a compilation that contains two or more individually titled parts.

Collocation. The bringing together of metadata descriptions or information resources that are related in some way (e.g., same author, same work, same subject).

Collocation device. A number or other designation on a resource used to place it next to (i.e., collocate with) other resources that are like it.

Colon Classification. Classification scheme devised by S. R. Ranganathan in the early 1930s; it was the first fully faceted classification scheme. Its name comes from Ranganathan's early use of the colon punctuation symbol to separate facets.

Colophon. A set of data at the end of a resource that gives varying kinds of bibliographic data. It might give information usually found on a title page, and, in items after the invention of printing with moveable type, it gives such information as date of printing, printer, typeface used, and the like.

Command searching. The process of searching in which a code (e.g., "a" or "au" for *author*, "t" or "ti" for *title*) is followed by an exact string of characters that is matched against the system's internal indexes.

Compilation. A collection of works (or parts of works) by one author published together, or a collection of two or more works or parts of works by more than one author published together. Each work in a compilation was originally written independently or as part of an independent publication.

Comprehensive description. A description of a whole resource (as opposed to describing only a part). *See also* **Analytical description; Hierarchical description**.

Computer output microform (COM) catalog. A catalog that is produced on either microfiche or microfilm and that requires a microform reader for its use.

Conceptual analysis. An examination of the intellectual or creative contents of an information resource to understand what the item is about and what the item is (i.e., its form or genre). *See also* **Subject analysis; Translation.**

Container architecture. A model for pulling together distinct packages of metadata, which are related to the same information resource, into a single container. This conceptual model allows different communities to create, maintain, and share their metadata.

Container list. *See* **Finding aid.**

Content. The intellectual information transmitted in or by an information resource or metadata; content is distinguished from the encoding, packaging, or framework used for transmission.

Content standard. A set of rules or instructions to guide catalogers, indexers, and the like, in the creation and formatting of data for a bibliographic or index record, an authority record, a metadata statement, or some other form of resource description. *See also* **Cataloging code**.

Content type. In RDA, a category representing how the content of a work is communicated or perceived (e.g., *text*, *performed music*, *still image*, *cartographic three-dimensional form*).

Contextualize. A user task in FRAD; to place in context or to clarify relationships. This task has been deprecated in IFLA LRM.

Continuing resource. *See* **Integrating resource; Serial.**

Contract cataloging. An institution's use of a contractual relationship with a person or agency to provide metadata that represents the institution's acquisitions for its collection. *See also* **Outsourcing**.

Contributor. An agent that has made significant contributions, other than principal creation, to a resource; in RDA, an agent that helps to realize an expression of a work—sometimes in the original creation of the work and sometimes in later expressions of the work—with supplementary content (such as illustrated editions or versions with commentary) that was not present in the original expression.

Control field. A field in the MARC format (00X) that includes numeric or other encoded data for retrieval.

Controlled access point. A name, term, or code found in an authority record; may be an authorized access point or a reference. *See also* **Access point; Authorized access point (AAP); Variant access point (VAP)**.

Controlled title. A title that has been established as part of authority work; it is used for work and expression titles (including series titles). *See also* **Uncontrolled title**.

Controlled vocabulary. A list or database of terms in which all terms or phrases representing a concept are brought together. Often a preferred term or phrase is designated for use in resource descriptions in a retrieval tool; the non-preferred terms have references from them to the chosen term or phrase, and relationships (e.g., broader terms, narrower terms, related terms) among preferred terms are identified. There may also be scope notes to explain the terms.

Conventional collective title. In RDA, a collective title, referring to the form of the work, used for a compilation containing two or more works of that type (e.g., *Plays*, *Works*, *Speeches*), or it may be used for two or more parts of a work (e.g., *Selections*).

Cooperative cataloging. The working together of independent institutions to create cataloging that can be shared with others.

Coordinate terms. Terms that share the same parent terms (e.g., *green* and *blue* are coordinate terms; they are both child terms of *primary colors*).

Copy cataloging. Adapting the original cataloging created by one library for use in another institution's catalog. *See also* **Original cataloging.**

Core element. An element in a metadata standard that is required in all descriptions (if the information is applicable and discernible).

Core record. *Obsolete.* A standard prescribed by the Program for Cooperative Cataloging (PCC) that presents the minimum requirements for elements to be included in a nationally acceptable AACR2 record.

Core-if element. In RDA, an element in a metadata standard that is required only in certain circumstances or under particular conditions.

Corporate body. An organization or a group of persons that is identified by a collective name and that acts (or may act) as an entity.

Crawler. *See* **Spider.**

Creator. An agent that is responsible for the intellectual or artistic content of a work; includes authors, writers, enacting jurisdictions, composers, photographers, artists, and the like.

Cross-reference. An instruction in a retrieval tool that directs a user to another place in the tool; also called a *reference*.

Crosswalk. A table, chart, or other device that indicates equivalence relationships among concepts, elements, or values in two or more controlled vocabularies, metadata schemes, encoding standards, and so on; for example, a crosswalk could be used to show which classification notation in DDC is equivalent to a notation in LCC.

Cutter number. A designation that has the purpose of alphabetizing all works that have exactly the same classification notation; named for Charles Ammi Cutter, who devised such a scheme (but used with a small *c* when referring to another such table that is not Cutter's own). *See also* **Call number.**

DACS. *See* ***Describing Archives: A Content Standard*** (DACS).

Data. Unprocessed information, which may be in the form of numbers (binary data, numerical data sets), text (facts, information without context), images, etc.

Data element identifier. A letter or number following a delimiter that specifies the type of content found in a MARC subfield (e.g., the letters in *$a* and *$b*). *See also* **Delimiter; Subfield code.**

Data value standard. A controlled list of terms or codes used to populate particular metadata elements. Examples include lists of resource types or subject headings. *See also* **Controlled vocabulary.**

Database. A set of records that are all constructed in the same way and are often connected by relationship links; the structure underlying retrieval tools.

Database index. A retrieval tool that provides access to the analyzed contents of information resources (e.g., articles in journals, short stories in collections, papers in conference proceedings, reviews, reports). A database index contains electronically accessible and searchable entries that provide descriptive and administrative metadata, as well as descriptor terms (and, in some instances, a classification notation) to represent aboutness. May be referred to as a *journal index* or a *periodical index. See also* **Index.**

DC. *See* **Dublin Core**.

DDC. See ***Dewey Decimal Classification*** **(DDC)**.

Deep web. The invisible, hidden parts of the web that are not indexed in search engines.

Delimiter. A special character or symbol that indicates the beginning of a particular MARC subfield (e.g., $ or | or ‡). *See also* **Data element identifier; Subfield code.**

Depth indexing. Assignment of all the subject terms necessary to represent all of the main concepts in a document, including many subtopics and underlying themes. *See also* **Exhaustivity; Summarization.**

Describing Archives: A Content Standard **(DACS)**. A content standard for archival description used in the United States. It may be used with EAD-encoded finding aids and MARC records for archival collections. It has replaced APPM. *See also* **APPM; Archival description.**

Description. That portion of the descriptive cataloging process in which elements that identify a resource are transcribed into a bibliographic record; also, the portion of the bibliographic record (i.e., descriptive data) that results from this process. *See also* **Access; Descriptive cataloging; Descriptive data; Metadata; Resource description.**

Descriptive cataloging. That phase of the cataloging process that is concerned with the identification and description of a resource, the recording of this information in a bibliographic record, and the selection and formation of access points—with the exception of subject access points. *See also* **Access; Description.**

Descriptive data. Data that describes a resource, such as its title, its edition, its date of publication, its extent, and notes identifying pertinent features.

Descriptive metadata. Metadata that contains the important identifying characteristics of a resource and the analysis of its contents for the purposes of discovery, identification, selection, and acquisition.

Descriptor. A subject term, representing a single concept, usually found in thesauri and used in indexes. *See also* **Subject heading**.

Designation of edition. A word or number identifying the edition to which a resource belongs.

Devised title. A title provided for a collection by an archivist or by another information professional for an untitled resource.

Dewey Decimal Classification **(DDC).** Classification devised by Melvil Dewey in 1876; it divides the world of knowledge hierarchically into 10 classes, which are in turn divided into 10 divisions, which are in turn divided into 10 sections, and so on with additional subdivisions, using the 10 digits of the Arabic numeral system (i.e., 0–9). DDC is enumerative but with many faceting capabilities, especially in its later editions.

Diacritic. Modifying mark over, under, or through a character to indicate that pronunciation is different from that of the character without the diacritic; also called *diacritical mark*.

Dictionary catalog. A catalog arranged or displayed in alphabetical order with records for names, titles, and subjects all intermixed. *See also* **Alphabetical catalog; Classified catalog; Divided catalog**.

Digital archives. *See* **Digital collection.**

Digital collection. A collection of information resources in digital form that are selected, brought together, organized, preserved, and to which access is provided over digital networks for a particular community of users. May be referred to as a *digital library*, *institutional repository*, or *digital archives*.

Digital library. *See* **Digital collection**.

Direct entry. A principle in the formulation of controlled vocabularies that stipulates the entry of a concept directly under the term that names it, rather than as a subdivision of a broader concept (e.g., Child rearing, *not* Children—Development and guidance).

Directory (MARC). A component of a MARC record containing a series of 12-character fixed-length segments that identify the field tag, length, and starting position of each data field in the record.

Directory (Web). An organized collection of links to websites.

Discovery interface. An application attached to a retrieval tool that allows for additional ways to browse and organize search results; often allows for faceted browsing. *See also* **Faceted browsing; Retrieval tool.**

Divided catalog. A catalog in which different types of records are arranged or presented in separate files or displays. In print catalogs, usually the subject entries are separated from other entries, and sometimes titles are also separated. Order is usually alphabetical in each section, but the subject section may be in classified order. Online catalogs are de facto divided catalogs when authors, titles, and subjects are searched and displayed separately. *See also* **Alphabetical catalog; Classified catalog; Dictionary catalog.**

Document. A resource; often associated in people's minds with text and illustrations having been produced on paper, but increasingly associated with other forms of information resources.

Document clustering. *See* **Clustering.**

Document retrieval vs. Information retrieval. A dichotomy that is created by the level of exhaustivity used in subject indexing; summarization allows for retrieval of a document which can, itself, then be searched for relevant information, whereas depth indexing allows for retrieval of information at a much more specific level than the whole document. *See also* **Depth indexing; Exhaustivity; Summarization.**

Domain. (1) A sphere of knowledge, influence, or activity. (2) In networking, a group of computers whose host names share a common name (i.e., the domain name).

DTD (Document type definition). An SGML or XML application; defines the structure of a particular type of document. *See also* **XML schema**.

Dublin Core (shortened form of Dublin Core Metadata Element Set). An internationally agreed-upon set of metadata elements, with associated attributes and properties, developed and maintained by the Dublin Core Metadata Initiative as a simple standard for resource description; the 15 elements are broad and generic and therefore can be used to describe a wide range of resources.

EAC–CPF (Encoded Archival Context–Corporate Bodies, Persons, Families). An XML schema to encode metadata about agents involved in the creation of archival materials or agents that are the subjects of archival materials. Among other things, it puts agents into historical, geographical, and chronological context.

EAD (Encoded Archival Description). An XML schema created specifically to encode finding aids; it provides the structure for an archival finding aid and defines its data components.

Edition. A particular version of a resource; a specific expression of the intellectual content of the work found in a resource. *See also* **Designation of edition.**

Electronic resource. A resource that requires the use of a computer to access its intellectual contents.

Element. An individual category or field that holds an individual piece of description of a resource; typical metadata elements include title, creator, creation date, subject, and the like.

Element refinement. In Dublin Core, a qualifier that (1) sharpens the focus of an element, that is, it makes the meaning of an element narrower or more specific, or (2) identifies a controlled vocabulary or a data type that aids in the interpretation of an element value.

Encoding. Using a record syntax or a coding scheme to make metadata electronically accessible. Encoding ensures that metadata is structured logically and that it may be communicated, shared, and displayed easily. Encoding entails the setting off of each part of a record (or each metadata statement) so that (1) each of the parts can be identified clearly; (2) the parts or statements may be displayed in certain positions according to the wishes of those creating a display mechanism; and (3) certain parts of a record can be searchable.

Entity. A thing (a person, subject, work, resource, etc.); in cataloging, entities are described by listing their various attributes; also known by the Latin word *res*.

Entity-relationship model. Data analysis method used to describe the requirements and assumptions in a system. These models comprise three components: entities are things about which information is sought, attributes are data collected about the things, and relationships provide structure for linking things within a system. FRBR and IFLA LRM are based on this type of modeling.

Entry. *Obsolete.* The place in a print retrieval tool where a surrogate record is found.

Entry vocabulary. Cross-references in a controlled vocabulary list that point users to authorized subject terms; terms not chosen as the authorized terms in a controlled vocabulary.

Entry word. The first word of an authorized access point.

Enumerative classification. A classification arrangement that attempts to assign a designation for every subject concept (simple or complex) required in the system. *See also* **Faceted classification; Hierarchical classification.**

Equivalence relationship. The relationship between synonymous or nearly synonymous terms; often represented by the thesaurus labels *USE* or *UF.*

ERIC thesaurus. A commonly used name for the *Thesaurus of ERIC Descriptors,* a thesaurus for indexing and searching documents indexed by the Educational Resources Information Center.

Exception (RDA). An instruction that describes a situation where the cataloger is expected to deviate from standard practice based on special circumstances or based on particular resource types.

Exemplar. A typical or standard model or example.

Exhaustivity. The number of concepts that will be considered in the process of providing subject analysis; the two basic degrees of exhaustivity are *depth indexing* and *summarization. See also* **Depth indexing; Summarization.**

Expansive Classification (EC). A late nineteenth-century classification scheme created by Charles Ammi Cutter in which a set of coordinated schedules gives successive development possibilities from very simple (broad) to very detailed (close) subdivision.

Explicit knowledge. Knowledge that is recorded, codified, or communicated in an overt form. *See also* **Tacit knowledge.**

Explore. A user task in IFLA LRM; to discover and to gain greater understanding of resources and entities.

Expression. The second level in describing a Group 1 entity in the FRBR conceptual model; an expression is the way that a work is communicated through alphanumeric characters, signs, images, movement, sounds, or the like; an intangible, abstract entity that is made tangible in a *manifestation.*

Extensibility (metadata). The ability to use additional metadata elements and qualifiers to adapt to the specific needs of a community or project. *See also* **Qualifier**.

Extent. The number and type of units and/or subunits that make up the carrier of a resource.

Extrinsic relationship. A relationship to another resource that is informative but not essential to understanding the resource. *See also* **Intrinsic relationship.**

Facet. A component (piece, side, or aspect) of a subject.

Facet indicator. A symbol, punctuation mark, or reserved digit signifying that the digits or letters following that symbol represent another aspect of the topic. For example, in *Colon Classification,* a period (full stop) is used to indicate the upcoming space facet (i.e., geographic characteristic) in classification notations.

Faceted browsing. A feature in discovery interfaces that allows the searcher to browse retrieved search results by various categories established in the metadata, such as resource type, date, form, subject, place, etc.

Faceted classification. A classification arrangement that has small notations standing for subparts of the whole topic, which, when strung together, usually in a prescribed sequence, create a complete classification notation for a multipart concept. *See also* **Enumerative classification; Hierarchical classification.**

Faceting. An approach to categorizing terms in a controlled vocabulary or an approach to organizing discrete concepts in a classification scheme so that terms/concepts with a similar function or a shared characteristic will be clustered together.

False drop. An irrelevant search result.

Federated searching. The ability to search and retrieve results from multiple sources of information while using only a single, common interface. It features an all-inclusive search box for multiple systems, which may include catalogs, indexes, databases, and other electronic resources. Also known as *meta-searching.*

Field. A separately designated part of an encoded record; it may contain one or more subfields.

Filing. The process of placing paper records (e.g., catalog cards, acquisition forms) in order, usually in drawers.

Find. A user task in FRBR and IFLA LRM; to search for entities that match specific criteria.

Finding aid. An inventory-like description of an archival collection; it describes the whole collection as well as groupings within the collection.

Fixed field. A field of an encoded record that is always the same length from record to record. In MARC, the 008 field is often referred to as *the* fixed field.

Fixed location. A set place where a physical information resource will always be found or to which it will be returned after having been removed for use. *See also* **Relative location.**

Flexibility (metadata). The ability to include as much or as little detail as needed in one's metadata.

Folksonomy. The aggregation of the tags created by a large number of individual users. The term is a blend of *folks* and *taxonomy. See also* **Tag (Web 2.0); Tagging**.

Form heading. *See* **Genre/Form heading**.

Four-fold path. In RDA, a model for recording metadata about a resource. It allows catalogers to choose one or more of the following approaches to description: unstructured descriptions, structured descriptions, identifiers, and URIs/IRIs. Not all paths are appropriate for every RDA metadata element.

FRAD. See ***Functional Requirements for Authority Data*** **(FRAD)**.

FRBR. See ***Functional Requirements for Bibliographic Records*** **(FRBR)**.

FRBR LRM. See *IFLA Library Reference Model* **(LRM).**

FRSAD. See *Functional Requirements for Subject Authority Data* **(FRSAD).**

Full entry. *See* **Main entry (record).**

Functional Requirements for Authority Data **(FRAD).** A conceptual model developed by IFLA that describes the need for and the uses of authority-controlled data by information professionals and information seekers; superseded by the *IFLA Library Reference Model*.

Functional Requirements for Bibliographic Records **(FRBR).** A conceptual entity-relationship model developed by IFLA that identifies various entities found in the bibliographic universe, attributes associated with those entities, relationships among the entities, uses of bibliographic data by information seekers, and the relationships between the attributes and the uses of bibliographic data; superseded by the *IFLA Library Reference Model*.

Functional Requirements for Subject Authority Data **(FRSAD).** A conceptual model developed by IFLA; designed as a companion to FRBR and FRAD, it focuses on subject authority data, particularly on *thema* and *nomen*; superseded by the *IFLA Library Reference Model*. *See also* **Nomen; Thema.**

Fundamental categories. *See* **PMEST.**

Fuzzy set theory of categorization. A theory that holds that some categories are not well defined and sometimes depend upon the observer, rather than upon a definition (e.g., people who are under five feet tall think the category *tall people* is larger than do people who are over six feet tall).

Gateway. A computer system or a web location that provides access to many different databases or other resources through the same interface. For subject gateways, *see* **Directory (Web).**

General International Standard Archival Description **(ISAD(G)).** An international content standard for archival description. It may be used as a national standard or as a basis for developing national standards. DACS, the standard for archival description in the United States, is ISAD(G) compliant. *See also* **Archival description**.

Genre/Form heading. A controlled vocabulary term that refers to the literary genre, artistic form, or the publication format of a work rather than to its topical content; sometimes said to reflect an item's *is-ness* when discussed in opposition to *aboutness*. *See also* **Is-ness.**

Genus–species relationship. In controlled vocabularies, a specific type of hierarchical relationship in which one term is a *type* or a *kind* (species) of the thing represented in the broader term (genus). For example, *pets* and *dogs* reflect a genus–species relationship because *dogs* are a type or kind of *pet*. Sometimes referred to as a *class–class member relationship*. *See also* **Broader term (BT); Hierarchical relationship; Narrower term (NT)**.

Gifts and exchanges. Gifts are resources that are donated to an institution. Exchanges are when institutions swap duplicate or unwanted items according to a mutually beneficial trade agreement with other institutions.

GMD (General Material Designation). *Obsolete*. In an AACR2 record, an indication of the class of item being described (e.g., art original, electronic resource, motion picture); replaced in RDA with three metadata elements: *carrier type*, *content type*, and *media type*. *See also* **Carrier type; Content type; Media type**.

Granularity. In metadata description, the level and depth at which information resources are described; in database design, a measure of the size or number of segments into which memory is divided.

GUI (Graphical User Interface). A computer interface that uses icons and other such graphics to make a screen more intuitive for users.

Hashtag. *See* **Tag (Web 2.0).**

Heading. *Obsolete* (in descriptive cataloging). (1) An access point printed at the top (head) of a copy of a surrogate record, or column of records, in a printed tool, or appearing at the top of a listing of related works in an online retrieval tool. (2) The exact string of characters of the authorized form of an access point as it appears in the authority record. Replaced in RDA with *authorized access point*. Still used in subject cataloging in the term *subject heading*. *See also* **Access point; Authorized access point (AAP); Subject heading.**

Hierarchical classification. A classification that attempts to arrange subjects in a series of ordered groups, some of which are subordinate to others—proceeding from broad classes to more specific subdivisions. *See also* **Enumerative classification; Faceted classification**.

Hierarchical description. A description of two levels of a resource: a *comprehensive description* of the whole resource with *analytical descriptions* of one or more of its parts; also called *multilevel description*; generally not used in U.S. cataloging but is more common in describing archival collections. *See also* **Analytical description; Comprehensive description**.

Hierarchical relationship. In controlled vocabularies, a relationship in which one term is designated as being subordinate to (or narrower than) another term; identified by the relationship designators *BT* and *NT*. Hierarchical relationships may encompass a *genus–species*, *instance*, or *whole–part relationship*. *See also* **Broader term (BT); Genus–species relationship; Instance relationship; Narrower term (NT); Whole–part relationship**.

Hierarchy. An arrangement by which categories are grouped in such a way that a concept (e.g., class or discipline) is subdivided into subconcepts that are equal in level of specificity to each other, each of those subconcepts are further subdivided, and so on.

Holdings. The resources that an information institution owns or to which it can provide access (e.g., copies of print or digital books, issues of a journal, DVD sets).

Holonym. A term that represents the whole, which is made up of parts (e.g., *face* is a holonym for the term *nose*—a part, or *meronym*. *See also* **Meronym.**

Homograph. One of two or more words that look the same but have different meanings.

Homophone. One of two or more words that are spelled differently but are pronounced the same way.

HTML (Hypertext Markup Language). A scheme for encoding text, pictures, and the like, so that they can be displayed using various programs because the coding is totally made up of ASCII text.

HTTP (Hypertext Transfer Protocol). The part of a URL that lets the system's browser know that a web page is being sought; the protocol itself defines how messages are formatted and is the protocol most often used to transfer information from World Wide Web servers to system browsers.

Hyperlink. An electronic connection between two separate pieces of information: it may be between two web pages, between two parts of an electronic resource, between text and an image, and so forth.

Hypernym. A term with a broad meaning under which more specific terms fall, following a genus–species relationship; it is a parent term or broader term (e.g., *fruit* is a hypernym and *strawberry* is one of its hyponyms). *See also* **Broader term (BT); Hyponym**.

Hypertext. Documents in which words, pictures, references, and the like may be linked to other locations or documents so that clicking on one of the links takes a person to related information.

Hyponym. A term that is more specific in meaning, which is placed under terms with broader meanings; it is a child term or narrower term (e.g., *strawberry* is a hyponym and *fruit* is its hypernym). *See also* **Hypernym; Narrower term (NT)**.

ICP. See ***Statement of International Cataloguing Principles.***

Identifier. A unique character string (e.g., an ISBN, URI, a record number) associated with an entity (e.g., a person, a resource, a concept) that differentiates it from other entities.

Identify. A user task in FRBR and IFLA LRM; to confirm that an entity corresponds to the one sought; to recognize a specific resource, agent, and so on.

IFLA (International Federation of Library Associations and Institutions). International organization for the promotion of library standards and the sharing of ideas and research.

***IFLA Library Reference Model* (LRM).** A conceptual model of the bibliographic universe; an entity-relationship model designed to harmonize and replace FRBR, FRAD, and FRSAD.

ILL (Interlibrary loan). The process of acquiring a physical resource or a copy of it from a library that owns it by a library that does not own it, usually for the purpose of re-lending it to a patron.

ILS (Integrated library system). Computer system that includes various modules to perform different functions while sharing access to the same database. *See also* **Library services platform**.

Imprint. The information in a textual publication that tells where it was published, who published it, and when it was published.

Index. A bibliographic tool that provides access to the analyzed contents of resources (e.g., articles in a journal, short stories in a collection, papers in a conference proceeding). A back-of-the-book index provides access to the analyzed contents of one work.

Indexer. A person who determines access points (usually subject terms, but also sometimes authors or titles) that are needed in order to make metadata, often for smaller units of information (e.g., articles, papers), available to searchers. Indexers often are employed by for-profit organizations. *See also* **Cataloger.**

Indexing. The process of creating metadata, especially the access points, for information resources, often for smaller units of information (e.g., articles, papers).

Indexing vocabulary. *See* **Controlled vocabulary.**

Indicators. In the MARC encoding standards, indicators for a field contain coded information that is needed for interpreting or supplementing data in the field.

Individual (Ontologies). *See* **Entity**.

Infoglut. *See* **Information overload.**

Information. The communication or reception of knowledge; organized data. *See also* **Data; Knowledge.**

Information architecture. A methodology for planning, designing, building, organizing, and maintaining an information system (usually associated with systems on the web).

Information fragmentation. The situation that exists when information is scattered or spread across multiple devices or is found in different formats, such as when people have some documents or information kept on a smartphone, but other information is found on a desktop computer, a laptop, a tablet, and in other locations.

Information organization. The process of describing resources and then providing name, title, and subject access to the descriptions, resulting in resource descriptions that serve as surrogates for the actual items of recorded information and in resources that are logically arranged. Also referred to as *Bibliographic control* or *Organization of information.*

Information overload. When a person receives more information than can be processed or handled.

Information resource. *See* **Resource.**

Information retrieval. The process of gaining access to stored data for the purpose of becoming informed.

Information retrieval vs. Document retrieval. *See* **Document retrieval vs. Information retrieval**.

Information system. Technology used in information organization that addresses three basic functions: *storage*, *retrieval*, and *display*; may also be referred to as a *retrieval system* or an *information retrieval system*.

Initialism. An abbreviation made up of initial letters of words in a phrase, but each letter is pronounced separately (e.g., ABC is an initialism for the American Broadcasting Company). *See also* **Abbreviation; Acronym**.

Instance relationship. In controlled vocabularies, a specific type of hierarchical relationship in which one term is an example of another term. For example, *oceans* and *Indian Ocean* represent an instance relationship because *Indian Ocean* is an example of an *ocean*. *See also* **Broader term (BT); Hierarchical relationship; Narrower term (NT)**.

Institutional repository. *See* **Digital collection**.

Instruction sheet. In the *Subject Headings Manual* (SHM), a memo explaining policies of the Library of Congress for the correct application of and the establishment of terms for *Library of Congress Subject Headings* (LCSH).

Integrated library system or Integrated library management system. *See* **ILS (Integrated library system)**.

Integrating resource. A bibliographic resource that is added to or changed by means of updates that are integrated into the whole resource (includes updating loose-leafs and updating websites).

Interface design. The part of a system design that controls the interaction between the computer and the user.

Interlibrary loan. *See* **ILL.**

***International Standard Archival Authority Record for Corporate Bodies, Persons, and Families* (ISAAR(CPF)).** A content standard that provides guidance on creating archival authority records for corporate bodies, persons, and families who have created and maintained the records, documents, and other resources that comprise archives.

Interoperability. The compatibility of two or more systems such that they can exchange information and data and can use the exchanged information and data without any special manipulation.

Intrinsic relationship. A relationship that is essential for the identification of a work being cataloged. *See also* **Extrinsic relationship.**

Inventory. A tool whose purpose is to provide a record of what is owned.

IRI (Internationalized Resource Identifier). On the web, a means to uniquely identify entities such as people, corporations, books, abstract concepts, or network-accessible things via a character string, an address, a name, a number, or another such device; not limited to things that have network locations; uses Unicode as its character set. *See also* **URI (Uniform Resource Identifier).**

ISAAR(CPF). See ***International Standard Archival Authority Record for Corporate Bodies, Persons, and Families.***

ISAD(G). See ***General International Standard Archival Description.***

ISBD *(International Standard Bibliographic Description).* A standard that was designed in the early 1970s to facilitate the international exchange of cataloging records by standardizing the elements to be used in the description, assigning an order to those elements, and specifying a system of symbols to be used in punctuating the elements.

ISBN (International Standard Book Number). An internationally distinctive and unique number assigned to a monographic item.

ISM/LIS (Information Systems Management/Library Information Services). *See* **Auto-Graphics**.

Is-ness. The form or genre of a work; what the work is, rather than what it is about. *See also* **Aboutness; Genre/Form headings; Of-ness.**

ISO (International Organization for Standardization). A corporate body that oversees the creation and approval of standards.

ISSN (International Standard Serial Number). An internationally distinctive and unique number assigned to a serial.

Item. One copy of a manifestation of a work, such as a book, a map, an electronic file, or a sound recording, as distinct from its intellectual content (i.e., the work or expression of the work that it contains); the fourth level in describing a resource in the FRBR conceptual model.

Journal index. *See* **Database index.**

Justify. A user task in FRAD; to document the choice of name of, or inclusion of other attributes about, an entity (person, family, subject, etc.) and the reasons for the choices made. This task has been deprecated in IFLA LRM.

Keyword. A term that is chosen, either from actual text or from a searcher's mind, that is considered to be a *key* to finding certain information.

Keyword indexing. Use of significant words from a title or a text as index entries.

Keyword matching. A type of querying in retrieval systems where the computer matches discrete words, rather than exact phrases, to the data stored in the system's indexes. If more than one word is searched at the same time, the words do not need to be stored in the same index. *See also* **Phrase matching.**

Keyword searching. The use of one or more keywords as the intellectual content of a search command.

Knowledge. What exists in the mind (rather than in any stored form) of an individual who has studied a subject, understands it, and perhaps has added to it through research or other means; a combination of information, context, and experience.

Knowledge management. The attempt to capture, evaluate, store, and reuse what the employees of an organization know.

Knowledge organization system (KOS). A generic term for all types of schemes for organizing information, including classification schemes, categories, authority files, subject heading lists, thesauri, and ontologies.

Knowledgebase. A set of multiple different products, traditionally housed in separate databases but functionally integrated for the purposes of library resource management and discovery. Knowledgebases underlying *library services platforms* (LSPs) go beyond the functions typically included in *integrated library systems* (e.g., cataloging, authority control, circulation information) to include the offerings of library materials vendors (e.g., descriptions of accessibility of e-journals and e-books, URL link resolvers, discovery tools), dynamically updated within the LSP. *See also* **Library services platform (LSP).**

Known-item searching. Searching for a specific name, title, or subject within a retrieval tool to retrieve a particular information resource.

LC/NACO Authority File (LCNAF). A file housed at the Library of Congress containing not only the authority records created by LC and PCC contributors but also records contributed from Australia, Canada, Great Britain, and others; also called the *Library of Congress Name Authority File*.

LC-PCC PSs. See ***Library of Congress–Program for Cooperative Cataloging Policy Statements*** **(LC-PCC PSs).**

LCRIs. See ***Library of Congress Rule Interpretations.***

LCSH. See *Library of Congress Subject Headings*.

Leader. A component of a MARC record that identifies the beginning of a new record and provides coded information for record processing. The leader is fixed in length and contains 24 characters.

Left-anchored searching. A search in a retrieval tool in which the exact phrase is matched against data from the index starting with the first letter of the first word (or text string) moving from left to right letter-by-letter. Each letter in the query must exactly match the indexed data in order for it to be retrieved. *See also* **Phrase matching**.

Letter-by-letter filing. An arrangement of entries in a retrieval tool in which spaces and some punctuation marks are ignored so that an entry files as if it is all run together into one word (e.g., "New York" is treated as "Newyork" and follows "Newark"). *See also* **Word-by-word filing**.

Library guide. A subject bibliography that leads users to the resources a library has on a specific topic; may be in print or online; also known as a *research guide* or *pathfinder*.

Library housekeeping system. *See* **ILS (Integrated library system)**.

Library management support system. *See* **ILS (Integrated library system).**

***Library of Congress Classification* (LCC).** Classification scheme created by the Library of Congress beginning in the late 1890s; it divides the world of knowledge hierarchically into categories using letters of the English alphabet and then using Arabic numerals for further subdivisions. LCC is basically an enumerative scheme, allowing only a limited amount of faceting.

***Library of Congress–Program for Cooperative Cataloging Policy Statements* (LC-PCC PSs).** Interpretations or decisions—that have been made by LC's Policy and Standards Division, the PCC, or both—as to how catalogers will interpret, apply, or supplement specific RDA instructions.

***Library of Congress Rule Interpretations* (LCRIs).** *Obsolete.* A collection of the decisions that were made by the Library of Congress's Cataloging Policy and Support Office as to how catalogers at LC would interpret and apply AACR2; replaced by the *Library of Congress–Program for Cooperative Cataloging Policy Statements*.

***Library of Congress Subject Headings* (LCSH).** A list of terms to be used as controlled vocabulary for subject concepts by the Library of Congress and by any other agency that wishes to provide such controlled subject access to their metadata descriptions.

Library services platform (LSP). An extension of traditional integrated library systems, using current technology to address barriers to efficient use of divergent material types, particularly electronic resources. Typical characteristics include unified management of physical and electronic materials; use of global knowledgebases in addition to, or rather than, local databases; cloud computing as the basis for system architecture; and development of application programming interfaces to facilitate interaction of LSP system software with that of external vendors. *See also* **ILS (Integrated library system)**.

Linked data. A method of encoding and publishing data on the web, so that a wide range of different resources can be understood by computers as being related to the same entity or concept (e.g., if two documents are about Abraham Lincoln, they both can be identified as being about Lincoln, and they can be linked through automated means with the metadata used to identify the subject matter). Linked data makes possible the discovery of knowledge about entities that would otherwise have been separated by disassociated means of encoding or by different data silos. Linked data is referred to as open if it is made freely available with minimal or no restrictions on access or re-use. *See also* **Semantic Web**.

Literal. In metadata, a string of alphanumeric characters (e.g., "French" or "1985") providing the value for an attribute (as opposed to a URI).

Literary warrant. The principle that new notations are created for a classification scheme and new terms are added to a controlled vocabulary only when information resources actually exist about new concepts.

Location device. A number or other designation on an item to tell where it is physically located.

MADS (Metadata Authority Description Schema). An XML schema for an authority data element set that has been particularly developed for library applications; designed as a companion to MODS. *See also* **MODS (Metadata Object Description Schema).**

Main entry (access point). *Obsolete*. An access point that is chosen as the main or primary one; may also be referred to as *primary access point* in the library and archival worlds.

Main entry (record). *Obsolete*. A full catalog record, headed by the primary access point, that gives all the elements necessary for the complete identification of a manifestation of a work. This record also bears the tracings for all the other headings under which the work is entered.

Manifestation. The physical form in which an expression of a work can be found; the third level in describing a resource in the FRBR conceptual model.

Manuscripts. Papers created by an individual (not papers of a corporate body); original handwritten, typed, or word-processed documents that usually exist in single copies (unless they have been carbon-copied, photocopied, or printed multiple times).

MARC (MAchine-Readable Cataloging). A standard that prescribes codes that precede and identify specific elements of a bibliographic, authority, or holdings record, allowing the record to be read by a machine, which then displays the data in a fashion designed to make the record intelligible to users.

MARC 21. A MARC standard agreed upon by Canadian and U.S. representatives ("21" stands for the twenty-first century). MARC 21, which represents a consolidation of USMARC and CAN/MARC, two previous national MARC schemes, also has been adopted by Great Britain, Germany, and other countries, as well.

MARC record. An electronic bibliographic record that has its content designated according to MARC conventions.

MARC tag. A number that designates the kind of field in a MARC record.

Markup language. A scheme that allows the tagging and describing of individual structural elements of text for the purpose of digital storage, appropriate layout display, and retrieval of individual components (e.g., HTML, MARC, XML).

Media type. A category indicating the type of intermediation device (e.g., a computer, a projector) needed in order to interact with a resource; there are eight possible categories (i.e., *audio*, *computer*, *microform*, *microscopic*, *projected*, *stereographic*, *unmediated*, and *video*) as well as *other* and *unspecified*.

***Medical Subject Headings* (MeSH).** A list of terms created by the National Library of Medicine to be used as controlled vocabulary for subject concepts in the field of medicine.

Menu searching. A process of searching that allows one to navigate by making choices, not by giving commands.

Meronym. A term that represents one part of a whole (e.g., *nose* is a meronym of *face* [the holonym]). *See also* **Holonym.**

Meta tag. A tag in the header of an HTML document that contains metadata.

Metadata. Structured information that describes the attributes of resources for the purposes of identification, discovery, selection, use, access, and management; an encoded description of a resource (e.g., an RDA record encoded with MARC, a Dublin Core record); the purpose of metadata is to provide a level of data at which choices can be made as to which resources one wishes to view without having to search through massive amounts of irrelevant full text. *See also* **Bibliographic data; Description; Resource description; Surrogate record**.

Metadata registry. A database used to organize, store, manage, and share metadata schemas and their components.

Metadata statement. An assertion about a single attribute or property of a resource (e.g., an RDF triple indicating the title of a resource); metadata statements are the foundation of bibliographic or metadata records.

Meta-language. A set of rules for designing markup languages.

Meta-metadata. Data describing metadata (e.g., administrative data used to track metadata).

Meta-searching. *See* **Federated searching.**

METS (Metadata Encoding and Transmission Standard). A standard for encoding descriptive, administrative, and structural metadata for objects in digital collections.

Microdata. An HTML specification used to nest metadata annotations within the content of web resources; used by search engines and web crawlers to provide more relevant results for users.

Mode of issuance. A category reflecting how a resource is produced in terms of the number of parts, distribution, updating, and termination. *See also* **Integrating resource; Monograph; Multipart monograph; Serial**.

Model. An abbreviated description of a complex situation, paradigm, or process used to explain or demonstrate the situation, paradigm, or process.

MODS (Metadata Object Description Schema). A schema for a bibliographic element set that has been particularly developed for library applications; a subset of MARC expressed in XML; MADS, a set of authority data elements, has been established as a companion schema. *See also* **MADS (Metadata Authority Description Schema).**

Monograph. A complete bibliographic unit or information resource. It is often a single work but may also be one work or more than one work issued in successive parts; unlike serials, it is not intended to continue indefinitely.

Monographic series. *See* **Series (bibliographic)**.

Multilevel description. *See* **Hierarchical description.**

Multipart monograph. A monograph issued in a finite number of parts; it may be issued in successive parts at regular or irregular intervals, but it is not intended to continue indefinitely.

Museum accession record. A record used as a surrogate for an object acquired by a museum; it contains many kinds of information about the object, such as its provenance, financial history, location in the museum, historical significance, and the like.

Museum registration. *See* **Registration**.

Namespace. The authoritative place (i.e., domain or a directory) created to hold a collection of elements or attributes, each identified by a unique name or label (i.e., a URI); the collection (i.e., namespace) is also identified by a unique name or label (i.e., a URI).

Name/Title access point. An authorized access point that includes the authorized name of an agent and the preferred title for a work. It serves to identify a work.

Narrower term (NT). A term one level down from the term being considered in a listing where terms for subject concepts have been organized into relationships that are hierarchical. *See also* **Broader term; Related term.**

National Union Catalog **(NUC).** A publication of the Library of Congress that cumulated cataloging records from many libraries and indicated by NUC symbol those libraries that owned a particular resource; ceased publication in 2002.

Natural language. The language used by a person when expressing a concept about which information is desired.

Natural language processing (NLP). Computer analysis of written or spoken language in order to interpret meaning in a way that can allow the computer to understand and respond.

Navigate. In the ICP, a function of the catalog in addition to the ones articulated in FRBR and LRM. To navigate is to use the internal data structures of the information system to find sought information.

NISO (National Information Standards Organization). A corporate body that oversees the creation and approval of standards to be used in information processing; an American counterpart to ISO.

NLP (Natural language processing). *See* **Natural language processing.**

Nomen. In IFLA LRM, an entity's name or label.

Nonbook materials. Terminology used for any information resources that are not text in book form.

Non-literal. In metadata, an actual physical, digital, or conceptual entity, as opposed to a *literal* (i.e., an alphanumeric string representing an entity).

Notation. A representation in a system, such as a classification scheme, with a set of marks, usually consisting of letters, numbers, and/or symbols.

Number building. In DDC, the process of developing a complex classification number for a resource by appending digits from tables or other parts of the schedules to a base number.

Obtain. A user task in FRBR and IFLA LRM; to gain access to the resource described.

OCLC Connexion. The cataloging interface to OCLC's WorldCat database.

OCLC Online Computer Library Center. A bibliographic network, based in Dublin, Ohio, that is the largest and most comprehensive bibliographic network in the world.

Of-ness. What a visual resource is an image *of*, rather than what the work is (genre/form) or is about (subject matter). *See also* **Aboutness; Is-ness.**

One-to-one principle. In Dublin Core, the principle that descriptions should refer to only one resource at a time.

ONIX (Online Information Exchange). The book industry standard for representing and communicating product information in an electronic format; designed to carry the kind of information traditionally carried on jacket covers and, optionally, also to carry excerpts, book reviews, cover images, author photos, and the like.

ONIX message. Electronic information about a book, serial, or another resource transmitted by publishers and other participants in the book industry.

Online catalog. A catalog, available for use by the general public, in which resource descriptions are encoded for computer display; typically part of an integrated library system (ILS); also referred to as an *online public access catalog* or *OPAC*.

Online index. An index in which resource descriptions for the analyzed contents of information resources are encoded for computer display.

Ontology. A formal representation to machines of what, to a human, is common sense or reality; a formal naming of the entities that exist for a particular domain, with an attempt to define the types, properties, and relationships among them, which then define the essence of a situation, domain, or conceptual framework.

OPAC (Online public access catalog). *See* **Online catalog.**

Open stacks. The name given to the situation where patrons of a facility have the right to go into the storage areas themselves. *See also* **Closed stacks**.

Optional addition. An instruction in RDA that allows catalogers to supplement the information required by the preceding instruction.

Optional omission. An instruction in RDA that allows catalogers to omit some information called for in the preceding instruction.

Organization of information. *See* **Information organization.**

Organization of knowledge. *See* **Information organization.**

Organize. To perform the process of forming unity and arranging separate parts into a whole that functions as an integrated unit.

Original cataloging. The process of creating a bibliographic description "from scratch," especially without reference to other records for the same resource; also, the cataloging data created by this process. *See also* **Copy cataloging.**

Original order. In archival collections, the organization or sequence of records as established by the creator of those records; the archival order reproduces the order employed when the records were in active use.

Other title information. Words or phrases (e.g., a subtitle) that appear in conjunction with and are subordinate to the title proper of a resource; an additional title, phrase, or statement that helps to qualify or amplify the title proper (e.g., *A Moon for the Misbegotten: A Play in Four Acts*).

Outsourcing. A management technique used by an institution whereby some activities, formerly conducted in-house (e.g., acquisitions, cataloging, serials control, reference work), are contracted out for completion by a contracting agency. *See also* **Contract cataloging**.

Parallel title. A title proper that is repeated in another language or script.

Paris Principles. The conventional name of the Statement of Principles agreed upon by attendees at the International Conference on Cataloguing Principles in Paris, October 9–18, 1961; replaced by IFLA's *Statement of International Cataloguing Principles*.

Party. *See* **Agent.**

Pathfinder. *See* **Library guide.**

Patron-driven acquisitions. A process in which a library obtains certain resources only after the need for them has been definitively expressed by patrons; also known as *demand-driven acquisitions*.

PCC (Program for Cooperative Cataloging). An international cooperative program coordinated jointly by the Library of Congress and participants around the world; effort is aimed at expanding access to collections through useful, timely, cost-effective cataloging that meets internationally accepted standards.

Periodical. A publication with a distinctive title, which appears in successive numbers or parts at stated or regular intervals and which is intended to continue indefinitely. Usually each issue contains articles by several contributors. *See also* **Serial**.

Periodical index. *See* **Database index**.

Personal information management. The activities that individuals perform to organize, store, and retrieve information for their own purposes.

Phrase matching. A type of querying in retrieval systems where the computer matches exact phrases, rather than discrete words, to the data stored in the system's indexes. The exact phrase must be found with all the words together in the same index. *See also* **Keyword matching.**

Pick list. *See* **Simple term list**.

PMEST. A mnemonic device for identifying Ranganathan's five fundamental categories: *personality*, *matter*, *energy*, *space*, and *time*. The fundamental categories delineate five general facets that might be represented in any topic. *See also* **Faceted classification.**

Policy statement. Supplemental information used in coordination with RDA that provides an interpretation of an instruction, guidance on applying alternatives and options, or an additional policy that applies to an instruction.

Post-coordinate indexing. Indexing that enters subject concepts as single concepts so that searchers are required to coordinate them using such techniques as Boolean searching in order to locate resources about the compound and/or complex subjects in which the searchers are interested.

Precision. The measurement of how many of the documents retrieved are relevant. *See also* **Recall**.

Pre-coordinate indexing. The assigning of subject terms to surrogate records in such a way that some concepts, subconcepts, place names, time periods, and form concepts are put together in subject strings, and searchers of the system do not have to coordinate these particular terms themselves.

Preferred name. The name or form of name chosen to represent an agent; it is used as the basis for the AAP representing that agent.

Preferred sources of information. In RDA, the source or sources in an information resource considered to be the major location(s) for the bibliographic data to be used in preparing a description (e.g., title page of a book, title screen of a motion picture, label on a sound recording tape or disc); unlike the concept it replaced (i.e., *chief source of information* in AACR2), more than one location within the resource may be considered primary; it varies according to the type of resource.

Preferred term. A term or phrase chosen in a controlled vocabulary to be used consistently to represent a concept. Also referred to as an *authorized term*. *See also* **Descriptor; Subject heading**.

Preferred title for work. A standardized title chosen to identify a work that has been known by various titles over time or across multiple languages; formerly known as *uniform title*.

Preservation metadata. Metadata that supports and documents the preservation process used in libraries, archives, etc., to salvage damaged materials.

Primary access point. *See* **Main entry (access point).**

Printed catalog. A catalog in which the resource descriptions appear in static printed form (e.g., cards in a card catalog, columns in a book, or on microform).

Property. (1) In metadata, an element or attribute of a resource, such as the title, the description, or the creator of the resource. (2) In legal terms, it designates ownership or proprietorship, as in *intellectual property*.

Property-value pair. An underlying model for metadata statements; in order for metadata to be useful, each resource's attributes are described using properties (e.g., publication date) and their value (e.g., 2017). Paired together they communicate a statement about a resource. If either the value or the property is missing, the statement is incomplete.

Protocol. A standard set of rules that determines how computers communicate with each other across networks (e.g., HTTP and Z39.50 are protocols); it describes the format that a message must take and the way in which computers must exchange a message.

Prototype theory of categorization. The theory that categories have prototypes (i.e., best examples); for example, most people think a robin is a better example of a bird than is an ostrich.

Provenance. The origin or ownership trail of an archival document or collection, or of a museum object (i.e., information about its origins, custody, or ownership).

Proximity (in searching). The concept of *nearness* of search terms to one another in a text.

Public services. A library's front-of-house operations (as opposed to *technical services*), which focus on serving patrons through reference services, library instruction, reserves, circulation, and so on.

Publisher. The agent (e.g., corporate body or person) responsible for issuing information resources to make them available for public use.

Qualifier. A refinement to a metadata element to either sharpen the focus of the element or to identify a controlled vocabulary from which the value has been supplied. *See also* **Extensibility (Metadata)**.

Querying. The act of requesting information from a retrieval tool; conducting a search. *See also* **Browsing.**

RDA: Resource Description & Access. A set of cataloging instructions, originally based on FRBR and FRAD (and later on IFLA LRM), published in 2010 and implemented in 2013, for producing the description and name and title access points representing a resource; the descriptive cataloging standard that replaced AACR2. The creation of RDA was the result of collaboration among representatives from Australia, Canada, Germany, Great Britain, and the United States.

RDA Toolkit. A subscription-based electronic tool that contains the text of RDA, AACR2, related policy statements, and other resources useful in descriptive cataloging.

RDF (Resource Description Framework). An infrastructure that enables the encoding, exchange, and reuse of structured metadata; it uses XML as the means for exchanging and processing the metadata. RDF is based on the premise that resources have properties, properties have values, some values can be other resources with their own properties and values, and all these relationships can be linked within the framework. The basic structure of RDF is the RDF triple. RDF is the foundation for linked data and the Semantic Web.

RDF dataset. A group of RDF graphs.

RDF graph. A set of RDF triples.

RDF triple. The model used to structure metadata statements. A triple consists of a subject (a resource), a predicate (a property), and an object (a value). This is the underlying structure of linked data and the Semantic Web.

Recall. The measurement of how many of the relevant documents in a system are actually retrieved. *See also* **Precision.**

Record. *See* **Bibliographic record; Metadata; Surrogate record.**

Records management. The process of maintaining records for an organization; it includes such functions as making decisions about what records should be created, saving necessary records, establishing effective systems for retrieval of records, and archiving important records for posterity.

Reference. *See* **Cross-reference**.

Reference database. A database that contains links or pointers to the actual information held outside the database. *See also* **Source database.**

Refinement. *See* **Element refinement; Qualifier**.

Register. One of the control tools for a museum; it functions like a catalog with a number of additional kinds of access points (e.g., donor, style, provenance).

Registration. The process of creating documentation that uniquely identifies an object belonging to a museum; the object descriptions form the register (or catalog) for the museum.

Related term (RT). A term at the same level of specificity or bearing a non-hierarchical relationship to another term in a listing where terms for subject concepts have been organized into relationships that are hierarchical; this relationship suggests that if the concept already located is more or less relevant, there are other associated concepts that might also yield relevant material. *See also* **Broader term; Narrower term.**

Relational database. A form of database architecture in which records are structured in such a way that information is not all stored in the same file; files for different kinds of information are created (e.g., a bibliographic file, a personal name file, a corporate name file, a subject file, a classification file); records in the bibliographic file contain pointers to records in the other files and vice versa. A relational database structure conserves storage space, allows for faster searching, and allows for easier modification of records. Pointers establish relationships among records.

Relationship. A reciprocal association, link, or connection between two entities.

Relationship designator. A device (i.e., a label, phrase, or term) that clearly identifies the specific nature of the relationship that exists between entities.

Relative location. The situation in which an information resource will be or might be in a different place each time it is reshelved because it is shelved in relation (usually classificatory) to entities already shelved. *See also* **Fixed location.**

Relevance. A measure of how pertinent retrieval results are to particular queries or user needs. This concept is defined and calculated quite differently among various information retrieval systems.

Relevancy ranking. An algorithmic method used by search engines and other retrieval tools to order the search results so that the records most likely to be of interest to users will be listed first.

Reprint. A new printing of an item either by photographic methods or by resetting unchanged text.

Res. The Latin word for *thing* or *entity*; used in the IFLA LRM model.

Research guide. *See* **Library guide.**

Resource. An instance of recorded information (e.g., book, article, video, web page, sound recording, electronic journal); *resource* is used in order to avoid using *book*, *DVD*, or other such specific designations; also called *document*, *information resource*, *library materials*, *object*, and so on.

Resource description. Full descriptive and access information for a resource; a set of metadata statements referring to the same resource; other terms used for descriptions of resources are *bibliographic description*, *metadata*, and *surrogate record. See also* **Bibliographic record; Metadata; Surrogate record.**

Resource Description and Access. See ***RDA: Resource Description & Access*.**

Resource discovery. The process of locating, accessing, retrieving, and bringing together relevant information from widely distributed networks.

***Respect des fonds*.** The principle that states that archival materials created or collected together should be kept together without mixing in records or materials from other creators or collections. *See also* **Original order; Provenance.**

Retrieval tool. A device such as a catalog, an index, a search engine, and the like, created for use as an information retrieval system.

Retrospective conversion. The process of changing information in eye-readable surrogate records into machine-readable form.

Rights and access metadata. Metadata addressing access, use of information resources, or intellectual property rights.

Robot (Internet). *See* **Spider**.

Romanization. The representation of the characters or script of a non-Roman alphabet by Roman characters.

Rule of three. *Obsolete.* In AACR2, a restriction placed on the number of access points for creators of and contributors to the resource—if more than three were involved, only the first of each kind was recorded; also, there was a restriction on the numbers of persons or corporate bodies listed in statements of responsibility—if more than three were found on the chief source of information, only the first was recorded and the others were replaced by the Latin abbreviation *et al.* (meaning *and others*). In RDA, no such restrictions apply.

Schedule. *See* **Classification schedule.**

Schema. (1) A set of metadata elements created by a particular community or for a particular type of resource. (2) A document or piece of code that controls a set of terms in another document or piece of code; similar in function to a master checklist.

Schema-neutral. The idea that a content standard, controlled vocabulary, or some other standard may be used with more than one metadata scheme, display format, or encoding format.

Scope note. A statement delimiting the meaning and application of a subject heading, index term, or classification notation.

Search engine. A computerized retrieval tool that, in general, matches keywords input by a user to words found in documents of the site being searched; the more sophisticated search engines may allow other than keyword searching.

Search engine optimization (SEO). An activity aimed at raising the visibility of websites by getting sites to be ranked higher in search engine results.

Sears List of Subject Headings (Sears). A controlled vocabulary of terms and phrases that is used mostly in small libraries to provide subject access to resources available from those libraries.

See also note (SA). A reference in LCSH instructing the cataloger to consider other ways to express the concept found in the main heading (e.g., as a subdivision, or in a phrase heading).

Select. A user task in FRBR and IFLA LRM; to choose a resource that is appropriate to the user's needs.

Selection (Collection development). The process in which collection development librarians learn about the existence of works through vendors' product catalogs, reviews, publishers' announcements, and the like, and then choose the most appropriate materials for the collection.

Semantic Web. An extension of the World Wide Web. The *traditional web* provides linkages between online resources, generally at the level of the whole resource or a discrete part of it. The *Semantic Web* provides linkages among statements about resources, in a format semantically meaningful to, and actionable by, computers. Linked data is generally considered to be the basis for the Semantic Web. *See also* **Linked data.**

Semantics. The meaning of a string of characters, as opposed to the syntax, which dictates the structure, independent of meaning.

Serial. A publication (physical or electronic) issued in successive parts (regularly or irregularly), and usually consecutively numbered, that is intended to continue indefinitely. Included are periodicals, newspapers, proceedings, reports, memoirs, annuals, numbered monographic series, and online journals. *See also* **Integrating resource.**

Series (archival). A logically grouped set of files, books, correspondence, or other such set.

Series (bibliographic). A number of separate works, commonly related in subject or form, that are issued successively and with some regularity. They are usually issued by the same publisher, distributor, etc., and are in uniform style, with a collective title. Each monograph in the series frequently is given a number, usually in chronological order. Works in a series *may* be cataloged together as a serial, especially if numbered.

SGML (Standard Generalized Markup Language). An international standard for document markup for machine readability.

Shelf reading. The process of scanning call numbers of shelved items to determine whether they are in correct order.

Shelflist. Originally, a list of physical information resources owned by an institution in the order in which they appeared on the shelves of the institution where they were housed; with time, the meaning has developed to indicate classification order display of surrogate records for information resources, which now allows for intangible as well as physical resources.

Shelving. The process of placing physical information resources on shelves in the order of the arrangement of their call numbers or other notations that indicate their appropriate locations.

Simple term list. A straightforward list of values that may be used for a particular metadata element, usually in alphabetical order (or some other logical arrangement).

Single unit. *See* **Monograph.**

SkyRiver. A computer-based bibliographic network offering its database and services to a variety of libraries.

Social indexing. *See* **Tagging.**

Social tagging. *See* **Tagging**.

Source database. A database that contains the actual information sought, rather than one that points to outside data sources. *See also* **Reference database.**

Specific entry. A principle observed in the application of controlled vocabularies, by which a cataloger or an indexer assigns to an information resource the most precise term available in the controlled vocabulary (or allowed to be created by the rules of the vocabulary), rather than assigning a broader heading. *See also* **Specificity.**

Specificity. The level of semantic depth that is addressed by a particular controlled vocabulary (e.g., LCSH has greater specificity in its established headings than does *Sears*). *See also* **Specific entry.**

Spider. An automated Internet program that gathers content from web pages (storing URLs and indexing keywords, links, and text) for inclusion in a search engine; may also be referred to as *crawler, web crawler, agent,* or *robot.*

SRU (Search/Retrieval via URL). A standard XML search protocol for Internet search queries that uses the Contextual Query Language (CQL), a standard syntax for representing queries.

Standard. Something established by authority or custom as a model or example; in the information field, standards are approved by national and/or international bodies after discussion and voting by representatives.

Statement of International Cataloguing Principles **(ICP).** An IFLA document that outlines principles and values associated with cataloging; designed as a replacement for the 1961 Paris Principles. *See also* **Paris Principles.**

Statement of responsibility. A transcription of a statement, from the preferred sources of information, naming the agents responsible for the intellectual or artistic content of a work, contributions to an expression, performances of the content, revisions of a work, subsequent editions, and so on.

Stereotyping. A method of printing using a metal copy of a typeset image.

Stopword list. A list of words *not* included in, or filtered out of, retrieval tools.

Structural metadata. Metadata addressing the "makeup" or structural composition of an electronic information resource.

Structure. (1) Arrangement in a definite pattern of the parts of a whole. (2) The data model used to shape the way that metadata statements are expressed.

Subdivision. (1) In controlled vocabulary, a term or phrase appended onto a subject heading to provide additional specificity, to show special treatment of a subject, or to bring out additional facets of the topic (e.g., geographic, chronological, form). (2) A level of structure in a hierarchical classification at which subordinate concepts are represented (the level below a class, division, section, etc.).

Subfield. A separately designated segment of a field in an encoded record.

Subfield code. A code, comprising a delimiter and an alphanumeric character, identifying the meaning of a particular segment of a field in an encoded record. *See also* **Data element identifier; Delimiter.**

Subject access. The provision to users of the means of locating information using subject terminology and/or classification notations.

Subject analysis. The part of indexing or cataloging that deals with the conceptual analysis of an information resource; the translation of that conceptual analysis into a framework for a particular classification, subject heading, or indexing system; and then using the framework to assign specific notations or terminology to the information resource and its surrogate record. *See also* **Conceptual analysis; Translation (Subject cataloging).**

Subject authority file. A record of choices made in the development of a controlled vocabulary. The authority file contains such things as justification for the choice of one synonym over another; references from unused synonyms or near-synonyms; references for broader terms, narrower terms, and related terms; scope notes; citations for references used; and the like. *See also* **Subject heading list.**

Subject cataloging. The process of providing subject analysis, including subject headings and classification notations, when creating resource descriptions for archives, libraries, museums, and the like.

Subject entry. *Obsolete.* The place in a retrieval tool where a surrogate record containing a particular controlled vocabulary term is found.

Subject gateway. *See* **Directory (Web).**

Subject guide. *See* **Library guide.**

Subject heading. Subject concept term or phrase found in a subject heading list and used in resource descriptions; sometimes used in indexes. *See also* **Descriptor.**

Subject heading list. A list of authorized controlled vocabulary terms or phrases together with any references, scope notes, and subdivisions associated with each term or phrase. *See also* **Subject authority file; Thesaurus.**

Subject indexing. The process of determining the aboutness of an information resource and determining appropriate subject terminology in an indexing language to express the aboutness in the resource description.

Subject language. A generic term for classification schemes and subject-focused controlled vocabularies; also called *indexing language*.

Subject subdivision. *See* **Subdivision.**

Subtitle. An additional title, subordinate to the title proper, that augments, expands, or limits the title proper; considered to be one kind of *other title information*.

Summarization. Indexing that identifies only a dominant, overall subject of an information resource, recognizing only concepts embodied in the main theme. *See also* **Depth indexing; Exhaustivity.**

Surrogate record. A presentation of the characteristics (e.g., title, creator, physical description if appropriate, date of creation, subject[s]) of a resource in one complete package (i.e., a *record*). *See also* **Bibliographic record; Metadata; Resource description.**

Switching language. A mediation language used to establish equivalencies between or among different subject indexing languages or classification schemes.

Synchronous responsibility. The situation in which all agents involved in the creation of an information resource have made the same kind of contribution to the creation of the work (e.g., joint authors). *See also* **Asynchronous responsibility.**

Syndetic structure. An organizational framework in which related names, topics, and other controlled terms are linked to each other via connective terms such as *See* and *See also*, or the thesaural indicators BT, NT, RT, USE, and UF.

Synonym. A term with the same meaning as another term; often, in controlled vocabularies, used for a term that has nearly the same meaning as well as for a term that has the same meaning. *See also* **Antonym.**

Synonym ring. A feature of a retrieval tool used to connect equivalent terms in the searching process (e.g., data tables, synonym lists).

Syntax. The arrangement of parts or elements so that they become constituents of a connected or orderly system; metadata's syntax is described by its encoding schema (e.g., MARC, XML), just as a language's syntax is described by its grammar.

System. *See* **Information system.**

System design. The specification of the working relations among all parts of a system; important design concepts include small testable components, user-friendly operation, and attractive functionality.

Table. *See* **Classification table**.

Tacit knowledge. Knowledge that is not recorded formally or is difficult to codify or share; it is knowledge still stored in the human mind. *See also* **Explicit knowledge**.

Tag (Encoding). A number, set of letters, certain set of punctuation marks, and so forth, that designates the kind of field in an encoding standard. *See also* **MARC tag**.

Tag (Web 2.0). In social media, the terms (keywords, subjects, etc.) assigned to a resource by a user; sometimes referred to as a *hashtag*.

Tagging. A populist approach to subject description. It is a process by which a distributed mass of users applies keywords to various types of web-based resources for the purposes of collaborative information organization and retrieval. Tagging allows individual users to group similar resources together by using their own terms or labels, with few or no restrictions. Also referred to as *user tagging, social tagging,* and *social indexing.*

Taxonomist. A person responsible for designing a controlled vocabulary or a classification scheme.

Taxonomy. A classification or controlled vocabulary, usually in a restricted subject field, that is arranged to show presumed natural relationships. *See also* **Classification scheme; Controlled vocabulary.**

Technical metadata. Metadata that describes the physical characteristics, origins, and lifecycles of digital resources; it is key to the preservation of the resource for future use. It gives the basic technical information that is needed in order to understand the nature of the information resource, the software and hardware environments in which it was created, and what is needed to make the resource accessible to users.

Technical services. The group of activities in an institution that involves acquiring, organizing, housing, maintaining, and conserving collections and automating these activities. In some places circulating collections is also considered to be a technical service.

TEI Header. A set of encoded metadata at the beginning of a TEI document that describes the document, its contents, and its origins.

Text Encoding Initiative (TEI). Refers to both the corporate organization with that name and to the encoding standard created by that group. The encoding standard was originally intended for the encoding of literary texts, although it has expanded to be used for other types of texts as well.

Thema. In FRSAD, it refers to the subject matter of a work; replaced in IFLA LRM by the entity *res*. *See also* **Res**.

Thesaurus. A specialized authority list of controlled vocabulary terms (usually restricted to a particular subject area) used with information retrieval systems; terms represent single concepts, together with any references, scope notes, and subdivisions associated with each term, and are organized so that the relationships between concepts are made explicit; similar to a list of subject headings, except for the emphasis on single terms rather than phrases. *See also* **Subject heading list**.

Time-span. In IFLA LRM, an entity reflecting a chronological period having a beginning, an end, and a duration.

Title. A name given to a resource. There are many types of titles. *See also* **Alternative title; Collective title; Conventional collective title; Devised title; Other title information; Parallel title; Subtitle; Title proper**.

Title entry. *Obsolete.* The place in a retrieval tool where a surrogate record containing the name of an information resource may be found.

Title proper. The main or primary title by which a resource is known; excludes any parallel title or other title information; includes alternative title and part title.

Tracing. *Obsolete.* On printed surrogate records (e.g., catalog cards, records in book catalogs), the set of name, title, and subject access points, other than the primary access point, that appear at the bottom of the record and are used to find (i.e., trace) the additional copies of the surrogate record.

Translation. The conveyance of the content of a work in another language. In RDA, a translation is considered a different expression of a work, not a new work.

Translation (Subject cataloging). The second stage of the subject analysis process, in which the aboutness is converted into terminology or symbols from one or more subject languages (e.g., controlled vocabularies, classification schemes). *See also* **Aboutness; Conceptual analysis; Subject analysis**.

Tree structure. A strict hierarchical arrangement with successive subdivisions; a diagram of which is said to resemble the branches of a tree.

Triple. *See* **RDF triple**.

Triplestore. A database specifically designed to house RDF structured data for the Semantic Web.

Turnkey system. A computer system customized to include all the hardware and software necessary for a particular function or application (e.g., a circulation system).

UBC. *See* **Universal Bibliographic Control.**

UCS (Universal Character Set). ISO standard for encoded representation of characters in computers; has the purpose of including all characters in all written languages of the world.

UDC. See ***Universal Decimal Classification.***

Uncontrolled access point. A name, title, and term that does not appear in an authority record; for example, a title proper is an access point but one that is transcribed and not standardized. *See also* **Authorized access point (AAP)**.

Uncontrolled title. A title that is transcribed from a manifestation; it is not an authority controlled title recorded in an authority record.

Unicode. An American industry counterpart to UCS, which permits computers to be able to handle the large number of character sets used in various languages. Both UCS and Unicode provide a unique number for every character to be used regardless of platform or format.

Uniform title. *Obsolete.* A title chosen for a work so that all manifestations will be displayed together under the same primary access point and also will be displayed together among all the entries for that access point. Uniform titles also are used to distinguish between and among different works that have the same title. The concept is replaced in RDA by *preferred title for the work*.

UNIMARC (Universal MARC). Originally conceived as a conversion format, in which capacity it requires that each national agency create a translator to change records from UNIMARC to the particular national format and vice versa; some countries have adopted it as their national format.

Union catalog. A catalog that represents the holdings of more than one institution or collection.

Universal Bibliographic Control (UBC). The concept that it will someday be possible to have access to metadata for all the world's important information resources.

***Universal Decimal Classification* (UDC).** A classification devised by Paul Otlet and Henri La Fontaine in the late 1890s. It was originally based on DDC, but has evolved into a much more faceted scheme than DDC.

Universal design. A strategy that extends system design to be as inclusive as possible (e.g., including ways for persons with hearing impairment or visual impairment or other disabilities to have the same access to digital information as do non-impaired persons).

URI (Uniform Resource Identifier). On the web, a means to uniquely identify entities such as people, corporations, books, abstract concepts, or network-accessible things via a character string, an address, a name, a number, or another such device; not limited to things that have network locations; unlike IRIs, URIs are limited to a subset of the ASCII character set, which limits their ability to display non-Roman scripts. *See also* **IRI (Internationalized Resource Identifier).**

URL (Uniform Resource Locator). One form of URI: the address of an information resource on the Internet; it indicates what protocol to use (e.g., *http*) and then gives the IP address or the domain name where the resource is located: most often server address, directory path, and file name.

User tagging. *See* **Tagging**.

User tasks. In several IFLA reports (i.e., FRAD, FRBR, FRSAD, ICP, and IFLA LRM), a set of activities that users hope to accomplish when interacting with information retrieval tools and bibliographic data. *See also* **Contextualize; Explore; Find; Identify; Justify; Navigate; Obtain; Select**.

USMARC (United States MARC). The version of MARC used in the United States until it was superseded by MARC 21.

Utlas International. *See* **Auto-Graphics.**

Value. A specific name of an attribute (or property) of a resource; for example, the value of a *language* attribute might be "French," or the value of a *duration* attribute might be "approximately 90 min."

VAP. *See* **Variant access point (VAP).**

Variable field. A field of an encoded record that can be as long or as short as the data to be placed into that field.

Variant access point (VAP). A string of alphanumeric characters, which represents an entity but is not authorized for use in resource descriptions; a variant access point refers users to an authorized access point. *See also* **Access point; Authorized access point (AAP); Controlled access point; Cross-reference**.

Variant name. A name or form of name *not* chosen to represent an agent; it is used as the basis for a cross-reference (i.e., a variant access point).

Vocabulary control. The process of creating and using a controlled vocabulary.

VRA (Visual Resources Association) Core. A set of guidelines that contains a metadata element set for describing visual documents depicting works of art, architecture, and artifacts or structures from material, popular, and folk culture.

Web. Short for World Wide Web (WWW); a nonlinear, multimedia, flexible system to provide information resources on the Internet and to gain access to such resources; based on hypertext and HTTP.

Web 2.0. The trend in web technology that emphasizes collaboration among users and interactivity between users and content. Examples of Web 2.0 features include rating resources, reviews, and tagging.

Web crawler. *See* **Spider**.

Whole–part relationship. In controlled vocabularies, a specific type of hierarchical relationship in which one term is a component of its broader term (i.e., the whole). For example, *nose* and *face* reflect a whole–part relationship because a *nose* is part of a *face*. *See also* **Broader term (BT); Hierarchical relationship; Holonym; Meronym; Narrower term (NT).**

Word-by-word filing. An arrangement of terms in a retrieval tool in such a way that spaces between words take precedence over any letter that may follow (e.g., "New York" appears before "Newark"). *See also* **Letter-by-letter filing.**

Work. A distinct intellectual or artistic creation; an abstract instance of content or ideas, regardless of the packaging in which the content or ideas may be expressed.

Work mark. A designation added to a cutter number, in a call number, that usually stands for the first word, not an article, of the title of the entity, but may stand for other entities, such as the name of a biographee, depending upon the circumstances.

WWW (World Wide Web). *See* **Web.**

XML (Extensible Markup Language). A subset of SGML, designed specifically for web documents, that omits some features of SGML and includes a few additional features (e.g., a method for reading non-ASCII text); it allows designers to create their own customized tags, thus overcoming many of the limitations of HTML.

XML schema. Like a document type definition (DTD), an XML schema provides definition and structure to XML documents. It outlines the constraints of XML documents and provides a mechanism for their validation by identifying approved elements and attributes, the number and order of elements, data types, and some values. Unlike a DTD, an XML schema is expressed in XML syntax itself, and follows XML rules. Moreover, XML schemas support namespaces, inheritance, and are capable of defining data types. *See also* **DTD (Document type definition)**.

Z39. The standards section of ANSI/NISO that is devoted to libraries, information science, and publishing.

Z39.50. A national standard that provides for the exchange of information, such as surrogate records or full text, between otherwise incompatible computer systems.

Z39.50 protocol. A standard applications-level tool that allows one computer to query another computer and transfer search results without the user having to know the search commands of the remote computer.

SELECTED BIBLIOGRAPHY

Agarwal, Naresh, and Md. Anwarul Islam. “Knowledge Management Implementation in a Library: Mapping Tools and Technologies to Phases of the KM Cycle.” *VINE: The Journal of Information and Knowledge Management Systems* 44, no. 3 (2014): 322–44.

Allen, Sharon. *Beginning Relational Data Modeling.* 2nd ed. Berkeley, CA: Apress, 2005.

Anglo-American Cataloguing Rules, Second Edition, 2002 Revision, prepared under the direction of the Joint Steering Committee for Revision of AACR. Ottawa: Canadian Library Association; Chicago: American Library Asso-cia-tion, 2002.

Antoniou, Grigoris, Paul Groth, Frank van Harmelen, and Rinke Hoekstra. *A Semantic Web Primer*. 3rd ed. Cambridge, MA: MIT Press, 2012.

Arms, William Y. *Digital Libraries.* Cambridge, MA: MIT Press, 2000.

Avram, Henriette. “Machine Readable Cataloging (MARC): 1961–1974.” In *Encyclopedia of Library and Information Science*, 3rd ed., edited by Marcia J. Bates and Mary Niles Maack. New York: Taylor and Francis, 2009.

Baca, Murtha, ed. *Introduction to Metadata*. 3rd ed. Los Angeles: Getty Research Institute, 2016. http://www.getty.edu/publications/intrometadata/.

Baca, Murtha. “A Picture Is Worth a Thousand Words: Metadata for Art Objects and Their Visual Surrogates.” In *Cataloging the Web: Metadata, AACR, and MARC 21,* edited by Wayne Jones et al., 131–38. Lanham, MD: Scarecrow, 2002.

Baca, Murtha, Patricia Harpring, Elisa Lanzi, Linda McRae, and Ann Whiteside. *Cataloging Cultural Objects: A Guide to Describing Cultural Works and Their Images.* Chicago: American Library Association, 2006. http://cco.vrafoundation.org/.

Baker, Nicholson. “Discards.” *New Yorker* 70, no. 7 (April 4, 1994): 64–86.

Banerjee, Sujata, and Vibhu O. Mittal. “On the Use of Linguistic Ontologies for Accessing and Indexing Distributed Digital Libraries.” In *Digital Libraries ’94: Proceedings of the First Annual Conference on the Theory and Practice of Digital Libraries, June 19–21, 1994, College Station, Texas*, edited by John L. Schnase. College Station: Texas A&M, 1994. http://www.csdl.tamu.edu/DL94/paper/banerjee.html.

Bastian, Jeannette, Megan Sniffin-Marinoff, and Donna Webber. *Archives in Libraries: What Librarians and Archivists Need to Know to Work Together*. Chicago: Society of American Archivists, 2015.

Bates, Marcia J. “The Design of Browsing and Berrypicking Techniques for the Online Search Interface.” *Online Review* 13, no. 5 (October 1989): 407–24.

Bates, Marcia J. “Subject Access in Online Catalogs: A Design Model.” *Journal of the American Society for Information Science* 37, no. 6 (1986): 357–76.

Bauer, Florian, and Martin Kaltenböck. *Linked Open Data: The Essentials*. Vienna: DGS, 2012.

Bennick, Ann, and Judy Vasek Sittig. *Managing Active Business Records*. Overland Park, KS: ARMA

Bennick, Ann, and Judy Vasek Sitton. *Managing Active Business Records*. Overland Park, KS: ARMA International, 2014.

Berman, Sanford. *Joy of Cataloging*. Phoenix, AZ: Oryx, 1981.

Berman, Sanford. *Prejudices and Antipathies: A Tract on the LC Subject Heads Concerning People*. Metuchen, NJ: Scarecrow, 1971.

Berner, Richard C. "Historical Development of Archival Theory and Practices in the United States." *Midwestern Archivist* 7, no. 2 (1982): 103–17.

Berners-Lee, Tim, with Mark Fischetti. *Weaving the Web: The Original Design and Ultimate Destiny of the World Wide Web by Its Inventor*. New York: HarperCollins, 1999.

Berners-Lee, Tim, James Hendler, and Ora Lassila. "The Semantic Web." *Scientific American* 284, no. 5 (May 2001): 34–38, 40–43.

Besson, Alain. *Medieval Classification and Cataloguing: Classification Practices and Cataloging Methods in France from the 12th to 15th Centuries*. Biggleswade, UK: Clover, 1980.

Bettington, Jackie, ed. *Keeping Archives*. 3rd ed. Canberra: Australia Society of Archivists, 2008.

Bierbaum, Esther Green. "Records and Access: Museum Registration and Library Cataloging." *Cataloging & Classification Quarterly* 9, no. 1 (1988): 97–111.

Borgman, Christine L. *From Gutenberg to the Global Information Infrastructure: Access to Information in the Networked World.* Cambridge, MA: MIT Press, 2000.

Borgman, Christine L. "Why Are Online Catalogs Hard to Use? Lessons Learned from Information Retrieval Studies." *Journal of the American Society for Information Science* 37, no. 6 (June 1986): 387–400.

Borgman, Christine L. "Why Are Online Catalogs *Still* Hard to Use?" *Journal of the American Society for Information Science* 47, no. 7 (July 1996): 493–503.

Bowker, Geoffrey C., and Susan Leigh Star. *Sorting Things Out: Classification and Its Consequences.* Cambridge, MA: MIT Press, 1999.

Breeding, Marshall. "Library Services Platforms: A Maturing Genre of Products." *Library Technology Reports* 51, no. 4 (2015): 1–40.

Breeding, Marshall. "Library Systems Report." *American Libraries.* May 1st issue each year.

Breeding, Marshall. *Library Technology Guides.* http://www.librarytechnology.org/.

Breeding, Marshall. *Next-gen Library Catalogs*. New York: Neal-Schuman, 2010.

Broughton, Vanda. *Essential Classification*. Chicago: ALA Neal-Schuman, 2015.

Broughton, Vanda. "The Need for a Faceted Classification as the Basis of All Methods of Information Retrieval." *Aslib Proceedings: New Information Perspectives* 58, no. 1/2 (2006): 49–72.

Brown, Caroline, ed. *Archives and Recordkeeping: Theory into Practice*. London: Facet, 2014.

Buck, Rebecca A., and Jean Allman Gilmore, eds. *MRM5: Museum Registration Methods*. 5th ed. Washington, DC: American Association of Museums, 2010.

Buckland, Michael K. "What Is a 'Document'?" *Journal of the American Society for Information Science* 48, no. 9 (September 1997): 804–9. http://people.ischool.berkeley.edu/~buckland/whatdoc.html.

Burke, Frank G. "Archives: Organization and Description." In *World Encyclopedia of Library and Information Services,* 3rd ed., edited by Robert Wedgeworth, 63–68. Chicago: American Library Association, 1993.

Bush, Vannevar. "As We May Think." *Atlantic Monthly* 176 (July 1945): 101–8. https://www.theatlantic.com/magazine/archive/1945/07/as-we-may-think/303881/.

Calhoun, Karen. *Exploring Digital Libraries: Foundations, Practice, Prospects*. Chicago: Neal-Schuman, 2014.

Caplan, Priscilla. *Metadata Fundamentals for All Librarians.* Chicago: American Library Association, 2003.

Carpenter, Michael. "The Original 73 Rules of the British Museum: A Preliminary Analysis." *Cataloging & Classification Quarterly* 35, no. 1/2 (2002): 23–36.

Casson, Lionel. *Libraries in the Ancient World.* New Haven, CT: Yale University Press, 2001.

Center, Clark E., Jr. *Records Management*. Washington, DC: Association of Research Libraries, 2008.

Chambers, Sally, ed. *Catalogue 2.0: The Future of the Library Catalogue*. Chicago: Neal-Schuman, 2013.

Chan, Lois Mai. *Library of Congress Subject Headings: Principles and Applications.* 4th ed. Westport, CT: Libraries Unlimited, 2006.

Chan, Lois Mai, Sheila S. Intner, and Jean Weihs. *A Guide to the Library of Congress Classification.* 6th ed. Santa Barbara, CA: Libraries Unlimited, 2016.

Chan, Lois Mai, Phyllis A. Richmond, and Elaine Svenonius, eds. *Theory of Subject Analysis: A Sourcebook*. Littleton, CO: Libraries Unlimited, 1985.

Chan, Lois Mai, and Athena Salaba. *Cataloging and Classification: An Introduction*. 4th ed. Lanham, MD: Rowman & Littlefield, 2016.

Clark, Alexander, Chris Fox, and Shalom Lappin, eds. *The Handbook of Computational Linguistics and Natural Language Processing*. Malden, MA: Wiley-Blackwell, 2010.

Classification Research Group. "The Need for a Faceted Classification as the Basis of All Methods of Information Retrieval." In *Theory of Subject Analysis: A Sourcebook,* edited by Lois Mai Chan, Phyllis A. Richmond, and Elaine Svenonius, 154–67. Littleton, CO: Libraries Unlimited, 1985.

Cleveland, Donald B., and Ana D. Cleveland. *Introduction to Indexing and Abstracting.* 4th ed. Santa Barbara, CA: Libraries Unlimited, 2013.

Collantes, Lourdes Y. "Agreement in Naming Objects and Concepts for Information Retrieval." PhD diss., Rutgers University, 1992.

Craven, Louise, ed. *What Are Archives? Cultural and Theoretical Perspectives: A Reader*. Burlington, VT: Ashgate, 2008.

Cutter, Charles A. *Rules for a Dictionary Catalog.* 4th ed. Washington, DC: Government Printing Office, 1904; Reprint, London: The Library Asso-cia-tion, 1962.

Dalkir, Kimiz. *Knowledge Management in Theory and Practice*. 2nd ed. Cambridge, MA: MIT Press, 2011.

Denton, William. "FRBR and the History of Cataloging." In *Understanding FRBR: What It Is and How It Will Affect Our Retrieval Tools,* edited by Arlene G. -Taylor, 35–57. Westport, CT: Libraries Unlimited, 2007.

Describing Archives: A Content Standard. 2nd ed. Chicago: Society of Amer-ican Archivists, 2013. http://www2.archivists.org/groups/technical-subcommittee-on-describing-archives-a-content-standard-dacs/dacs.

Desouza, Kevin C., and Scott Paquette. *Knowledge Management: An Introduction*. New York: Neal-Schuman, 2011.

DeWeese, Keith P., and Dan Segal. *Libraries and the Semantic Web*. San Rafael, CA: Morgan & Claypool, 2015.

Dillon, Andrew, and Don Turnbull. "Information Architecture." In *Encyclopedia of Library and Information Sciences*, 3rd ed., edited by Marcia J. Bates and Mary Niles Maack, 2361–68. New York: Taylor & Francis, 2009.

Ding, Wei, and Xia Ling. *Information Architecture: The Design and Integration of Information Spaces*. San Rafael, CA: Morgan & Claypool, 2010.

Dublin Core Metadata Initiative. http://dublincore.org/.

Dunkin, Paul S. *Cataloging U.S.A.* Chicago: American Library Association, 1969.

"EAD: Encoded Archival Description." Library of Congress. http://www.loc.gov/ead/.

EDItEUR. "ONIX: Overview." http://www.editeur.org/83/Overview/.

Enis, Matt. "All Systems Go: Library Systems Landscape 2016." *Library Journal* 140, no. 6 (2016): 34–38.

Enis, Matt. "Library Systems Landscape 2016: Company Profiles." *Library Journal* 140, no. 6 (2016): 42–44.

"ERIC's Indexing and Retrieval: 2001 Update" and "Thesaurus Construction and Format." In *Thesaurus of ERIC Descriptors,* 14th ed., edited by James E. Houston, xiv–xxxi. Phoenix, AZ: Oryx, 2001.

Evans, G. Edward, Sheila S. Intner, and Jean Weihs. *Introduction to Technical Services*. 8th ed. Santa Barbara, CA: Libraries Unlimited, 2011.

Farb, Sharon. "Universal Design and the Americans with Disabilities Act: Not All Systems Are Created Equal—How Systems Design Can Expand Information Access: 2000." In *From Catalog to Gateway: Charting a Course for Future Access: Briefings from the ALCTS Catalog Form and Function Committee,* edited by Bill Sleeman and Pamela Bluh, 99–106. Chicago: Association for Collections & Technical Services, American Library Association, 2005. http://www.pitt.edu/~agtaylor/ala/papers/FarbUniversalDesign.pdf.

Fishbein, Meyer H. "Archives: Records Management and Records Appraisal." In *World Encyclopedia of Library and Information Services,* 3rd ed., edited by Robert Wedgeworth, 60–63. Chicago: American Library Association, 1993.

Foskett, A. C. *The Subject Approach to Information.* 5th ed. London: Library Association Publishing, 1996.

Fox, Edward A., ed. *Digital Library Technologies: Complex Objects, Annotation, Ontologies, Classification, Extraction, and Security*. San Rafael, CA: Morgan & Claypool, 2014.

Furrie, Betty. *Understanding MARC Bibliographic: Machine-Readable Cataloging*. 8th ed. Washington, DC: Library of Congress, Cataloging Distribution Service, 2009. http://www.loc.gov/marc/umb.

Gallagher, Matt. "How to Conduct a Library Services Platform Review and Selection." *Computers in Libraries* 36, no. 8 (2016): 20–22.

The Getty. "Art & Architecture Thesaurus Online: About the AAT." http://www.getty.edu/research/tools/vocabularies/aat/about.html.

The Getty. "Getty Vocabularies." http://www.getty.edu/research/tools/vocabularies/index.html.

Glushko, Robert J., ed. *The Discipline of Organizing*. Cambridge, MA: MIT Press, 2013.

Godby, Carol Jean. *Mapping ONIX to MARC*. Dublin, OH: OCLC Research, 2010.

Gorman, Michael. *The Concise AACR2.* 4th ed. Chicago: American Library Association, 2004.

Gorman, Michael. "The Longer the Number, the Smaller the Spine; or, Up and Down with Melvil and Elsie." *American Libraries* 12, no. 8 (September 1981): 498–99.

Gross, Tina, and Arlene G. Taylor. "What Have We Got to Lose? The Effect of Controlled Vocabulary on Keyword Searching Results." *College & Research Libraries* 66, no. 3 (May 2005): 212–30.

Gross, Tina, Arlene G. Taylor, and Daniel N. Joudrey. "Still a Lot to Lose: The Role of Controlled Vocabulary in Keyword Searching." *Cataloging & Classification Quarterly* 53, no. 1 (2015): 1–39.

Guenther, Rebecca S. "MODS: The Metadata Object Description Schema." *Portal: Libraries and the Academy* 3, no. 1 (2003): 137–50.

Guerrini, Mauro, and Tiziana Possemato. "Linked Data: A New Alphabet for the Semantic Web." *Italian Journal of Library, Archives, and Information Science* 4, no. 1 (2013). https://www.jlis.it/article/view/6305.

Hagler, Ronald. *The Bibliographic Record and Information Technology.* 3rd ed. Chicago: American Library Association, 1997.

Harpring, Patricia. *Categories for the Description of Works of Art*, edited by Murtha Baca and Patricia Harpring. Revised March 2014 by Patricia Harpring. Los Angeles: The Getty, 2014. http://www.getty.edu/research/publications/electronic_publications/cdwa/.

Harris, Michael H. *History of Libraries in the Western World.* 4th ed. Lanham, MD: Scarecrow, 1995.

Hart, Amy. *RDA Made Simple: A Practical Guide to the New Cataloging Rules*. Santa Barbara, CA: Libraries Unlimited, 2014.

Heath, Tom, and Christian Bizer. *Linked Data: Evolving the Web into a Global Data Space*. San Rafael, CA: Morgan & Claypool, 2011.

Hedden, Heather. *The Accidental Taxonomist.* Medford, NJ: Information Today, 2010.

Hedden, Heather. *Indexing Specialties: Web Sites.* Medford, NJ: Information Today, 2007.

Hernandez, Michael J. *Database Design for Mere Mortals: A Hands-on Guide to Relational Database Design*. 3rd ed. Upper Saddle River, NJ: Addison-Wesley, 2013.

Hider, Philip. *Information Resource Description: Creating and Managing Metadata.* Chicago: ALA Editions, 2013.

Hider, Philip, with Ross Harvey. *Organising Knowledge in a Global Society: Principles and Practice in Libraries and Information Centres.* Rev. ed. Wagga Wagga, NSW: Centre for Information Studies, Charles Sturt University, 2008.

Hildreth, Charles R. "The Use and Understanding of Keyword Searching in a University Online Catalog." *Information Technology and Libraries* 16, no. 2 (June 1997): 52–62.

Hillmann, Diane, and Elaine L. Westbrooks, eds. *Metadata in Practice.* Chicago: American Library Association, 2004.

Hopkins, Judith. "The 1791 French Cataloging Code and the Origins of the Card Catalog." *Libraries & Culture* 27, no. 4 (Fall 1992): 378–404.

Hsieh-Yee, Ingrid. *Organizing Audiovisual and Electronic Resources for Access: A Cataloging Guide.* 2nd ed. Westport, CT: Libraries Unlimited, 2006.

Humbert, de Romans. *Regulations for the Operation of a Medieval Library.* St. Paul: Associates of the James Ford Bell Library, University of Minnesota, 1980.

Hunter, Eric J. *Classification Made Simple: An Introduction to Knowledge Organisation and Information Retrieval.* 3rd ed. Farnham, UK: Ashgate, 2009.

International Conference on Cataloguing Principles, Paris, 9th–18th October, 1961. *Report.* London, International Federation of Library Associations, 1963. Also available in: *Library Resources & Technical Services* 6 (1962): 162–67; and *Statement of Principles Adopted at the International Conference on Cataloguing Principles, Paris, October, 1961,* annotated ed., with commentary and examples by Eva Verona. London: IFLA Committee on Cataloguing, 1971; and A. H. Chaplin and Dorothy Anderson, eds. *Report/International Conference on Cataloguing Principles, Paris, 9th–18th October 1961.* London: IFLA International Office for UBC, 1981.

International Federation of Library Associations and Institutions, IFLA Study Group. *Functional Requirements for Bibliographic Records: Final Report.* Munich: Saur, 1998. https://www.ifla.org/publications/functional-requirements-for-bibliographic-records.

International Federation of Library Associations and Institutions, Meetings of Experts on an International Cataloguing Code. *Statement of International Cataloguing Principles*, 2016 edition. The Hague: IFLA, 2016. http://www.ifla.org/files/assets/cataloguing/icp/icp_2016-en.pdf.

International Federation of Library Associations and Institutions, Working Group on Functional Requirements and Numbering of Authority Records (FRANAR). *Functional Requirements for Authority Data: A Conceptual Model*, approved by the Standing Committees of the IFLA Cataloguing Section and IFLA Classification and Indexing Section, 2009, amended and corrected through July 2013. http://www.ifla.org/publications/functional-requirements-for-authority-data.

International Federation of Library Associations and Institutions, Working Group on the Functional Requirements for Subject Authority Records (FRSAR). *Functional Requirements for Subject Authority Data*: *A Conceptual Model.* Approved by the Standing Committee of the IFLA Section on Classification and Indexing, 2010. http://www.ifla.org/node/5849.

International Organization for Standardization. *Information and Documentation—Guidelines for Bibliographic References and Citations to Information Resources,* ISO 690:2010. 3rd ed. Geneva: ISO, 2010.

International Organization for Standardization. *Information and Documentation: Guidelines for the Content, Organization and Presentation of Indexes,* ISO 999:1996. Geneva: ISO, 1996.

International Standard Bibliographic Description (ISBD). Consolidated ed. Berlin: De Gruyter Saur, 2011. http://www.ifla.org/files/assets/cataloguing/isbd/isbd-cons_20110321.pdf.

Intner, Sheila, Susan S. Lazinger, and Jean Weihs. *Metadata and Its Impact on Libraries.* Westport, CT: Libraries Unlimited, 2006.

Intner, Sheila S., and Jean Weihs. *Standard Cataloging for School and Public Libraries*. 5th ed. Santa Barbara, CA: Libraries Unlimited, 2015.

ISAD(G): General International Standard Archival Description. Adopted by the Committee on Descriptive Standards, Stockholm, Sweden, 19–22 September, 1999. 2nd ed. Ottawa: International Council on Archives, 2000. http://www.ica.org/sites/default/files/CBPS_2000_Guidelines_ISAD(G)_Second-edition_EN.pdf.

Jackson, Sidney L. *Libraries and Librarianship in the West: A Brief History.* New York: McGraw-Hill, 1974.

Jacob, Elin K. "Classification and Categorization: A Difference That Makes a Difference." *Library Trends* 52, no. 3 (Winter 2004): 515–40.

Jacob, Elin K. "Ontologies and the Semantic Web." *Bulletin of the American Society for Information Science and Technology* 29, no. 4 (April/May 2003): 19–22. http://www.asis.org/Bulletin/Apr-03/BulletinAprMay03.pdf.

Jewett, Charles Coffin. *On the Construction of Catalogues of Libraries, and Their Publication by Means of Separate, Stereotyped Titles.* 2nd ed. Washington, DC: Smithsonian Institution, 1853.

Jin, Qiang. *Demystifying FRAD: Functional Requirements for Authority Data*. Santa Barbara, CA: Libraries Unlimited, 2012.

Johnson, Bruce Chr. "XML and MARC: Which Is 'Right'?" *Cataloging & Classification Quarterly* 32, no. 1 (2001): 81–90.

Jones, Catherine. *Institutional Repositories: Content and Culture in an Open Access Environment*. Oxford: Chandos, 2007.

Jones, William. *Keeping Found Things Found: The Study and Practice of Personal Information Management*. San Francisco, CA: Morgan Kaufmann, 2008.

Jones, William, and Jaime Teevan, eds. *Personal Information Management*. Seattle: University of Washington Press, 2007.

Joudrey, Daniel N. "Building Puzzles and Growing Pearls: A Qualitative Exploration of Determining Aboutness." PhD diss., University of Pittsburgh, 2005. http://d-scholarship.pitt.edu/id/eprint/10357.

Joudrey, Daniel N., Arlene G. Taylor, and David P. Miller. *Introduction to Cataloging and Classification*. 11th ed. Santa Barbara, CA: Libraries Unlimited, 2015.

Kao, Anne, and Stephen R. Poteet, eds. *Natural Language Processing and Text Mining*. London: Springer, 2007.

Keyser, Pierre de. *Indexing: From Thesauri to the Semantic Web*. Oxford: Chandos, 2012.

Knowlton, Steven A. "Three Decades since Prejudices and Antipathies: A Study of Changes in the Library of Congress Subject Headings." *Cataloging & Classification Quarterly* 40, no. 2 (2005): 123–45.

Kumar, Shiv. "From Clay Tablets to Web: Journey of Library Catalogue." *DESIDOC Journal of Library & Information Technology* 33, no. 1 (January 2013): 45–54.

Kwasnik, Barbara H. "How a Personal Document's Intended Use or Purpose Affects Its Classification in an Office." In *Proceedings of the 12th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, 207–10. New York: ACM, 1989.

Lai, Ling-Ling, and Arlene G. Taylor. "Knowledge Organization in Knowledge Management Systems of Global Consulting Firms." *Cataloging & Classification Quarterly* 49, no. 5 (2011): 387–407.

Lakoff, George. *Women, Fire, and Dangerous Things: What Categories Reveal about the Mind.* Chicago: University of Chicago Press, 1987.

Lancaster, F. W. *Indexing and Abstracting in Theory and Practice.* 3rd ed. Champaign: University of Illinois, Graduate School of Library and Information Science, 2003.

Lancaster, F. W. *Vocabulary Control for Information Retrieval.* 2nd ed. Arlington, VA: Information Resources, 1986.

Lancaster, F. W. "Whither Libraries? or Wither Libraries." *College & Research Libraries* 39, no. 5 (September 1978): 345–57; reprinted in *College & Research Libraries* 50, no. 4 (July 1989): 406–19.

Landis, William E., and Robin L. Chandler, eds. *Archives and the Digital Library*. Binghamton, NY: Haworth Information, 2006.

Langridge, D. W. *Classification: Its Kinds, Elements, Systems, and Applications.* London: Bowker-Saur, 1992.

Langridge, D. W. *Subject Analysis: Principles and Procedures.* London: Bowker-Saur, 1989.

Larson, Ray R. "Classification Clustering, Probabilistic Information Retrieval, and the Online Catalog." *Library Quarterly* 6, no. 2 (April 1991): 133–73.

Larson, Ray R., Jerome McDonough, Paul O'Leary, and Lucy Kuntz. "Cheshire II: Designing a Next-Generation Online Catalog." *Journal of the American Society for Information Science* 47, no. 7 (July 1996): 555–67.

Latham, Kiersten, and John E. Simmons. *Foundations of Museum Studies: Evolving Systems of Knowledge*. Santa Barbara, CA: Libraries Unlimited, 2014.

Leise, Fred. "Using Faceted Classification to Assist Indexing." *Key Words* 9, no. 6 (November/December 2001): 178–79.

Levy, David M. "Cataloging in the Digital Order." In *Digital Libraries '95: The Second Annual Conference on the Theory and Practice of Digital Libraries, June 11–13, 1995, Austin, Texas.* http://www.csdl.tamu.edu/DL95/papers/levy/levy.html.

Library of Congress. "Library of Congress Classification." https://www.loc.gov/catdir/cpso/lcc.html.

Library of Congress. "Manuscript Division Finding Aids Online." https://www.loc.gov/rr/mss/f-aids/mssfa.html.

Library of Congress, Catalogers Learning Workshop. "Library of Congress Subject Headings: Online Training." Janis L. Young and Daniel N. Joudrey, instructors. https://www.loc.gov/catworkshop/lcsh/.

Library of Congress, Network Development and MARC Standards Office. "BIBFRAME: Bibliographic Framework Initiative." http://www.loc.gov/bibframe/.

Library of Congress, Network Development and MARC Standards Office. "MARC Standards." http://www.loc.gov/marc/.

Library of Congress, Network Development and MARC Standards Office. "MARCXML: MARC 21 XML Schema." http://www.loc.gov/standards/marcxml/.

Library of Congress, Network Development and MARC Standards Office. "MODS: Metadata Object Description Schema." http://www.loc.gov/standards/mods/.

Library of Congress, Working Group on the Future of Bibliographic Control. "On the Record: Report of the Library of Congress Working Group on the Future of Bibliographic Control." http://www.loc.gov/bibliographic-future/news/lcwg-ontherecord-jan08-final.pdf.

Lubetzky, Seymour. *Cataloging Rules and Principles: A Critique of the A.L.A. Rules for Entry and a Proposed Design for their Revision.* Washington, DC: Library of Congress, 1953.

Malone, Thomas W. "How Do People Organize Their Desks? Implications for the Design of Office Information Systems." *ACM Trans Office Info Systems* 1, no. 1 (1983): 99–112.

Mann, Thomas. *The Oxford Guide to Library Research.* 4th ed. New York: Oxford University Press, 2015.

"The MARC 21 Formats: Background and Principles." [Prepared by] MARBI in conjunction with Network Development and MARC Standards Office, Library of Congress. Washington, DC: Library of Congress, 1996. http://lcweb.loc.gov/marc/96principl.html.

Marcella, Rita, and Arthur Maltby, eds. *The Future of Classification.* Aldershot, UK: Gower, 2000.

Markey, Karen. "The Online Library Catalog: Paradise Lost and Paradise Regained?" *D-Lib Magazine* 13, no. 1/2 (2007). http://www.dlib.org/dlib/january07/markey/01markey.html.

Maxwell, Robert L. *FRBR: A Guide for the Perplexed.* Chicago: American Library Association, 2008.

Maxwell, Robert L. *Maxwell's Handbook for RDA, Resource Description & Access: Explaining and Illustrating RDA: Resource Description and Access Using MARC21*. Chicago: American Library Association, 2013.

McIlwaine, Ia. *The Universal Decimal Classification: A Guide to Its Use*. The Hague: UDC Consortium, 2007.

McInerney, Claire R., and Michael E. D. Koenig. *Knowledge Management (KM) Processes in Organizations: Theoretical Foundations and Practice*. San Rafael, CA: Morgan & Claypool, 2011.

Millar, Laura A. *Archives: Principles and Practices*. New York: Neal-Schuman, 2010.

Miller, Steven J. *Metadata for Digital Collections: A How-to-Do-It Manual*. New York: Neal-Schuman, 2011.

Milsap, Larry. "A History of the Online Catalog in North America." In *Technical Services Management, 1965–1990: A Quarter Century of Change, A Look to the Future: A Festschrift for Kathryn Luther Henderson,* edited by Linda C. Smith and Ruth C. Carter, 79–91. Binghamton, NY: Haworth, 1996.

Mitchell, Erik T. *Library Linked Data: Early Activity and Development*. Chicago: ALA TechSource, 2016.

Mitchell, Erik T. *Metadata Standards and Web Services in Libraries, Archives, and Museums*. Santa Barbara: Libraries Unlimited, 2015.

Morville, Peter, and Jeffery Callender. *Search Patterns*. Sebastopol, CA: O'Reilly Media, 2010.

Moulaison, Heather Lea, and Raegan Wiechert. *Crash Course in Basic Cataloging with RDA*. Santa Barbara, CA: Libraries Unlimited, 2015.

Muller, Samuel, Johan A. Feith, and Robert Fruin. *Manual for the Arrangement and Description of Archives*. Translation of the second edition by Arthur H. Leavitt; with new introductions by Peter Horsman [and others]. Chicago: Society of American Archivists, 2003.

National Library of Medicine. "Welcome to Medical Subject Headings!" https://www.nlm.nih.gov/mesh/.

Neal, Diane Rasmussen, ed. *Indexing and Retrieval of Non-text Information*. Berlin: De Gruyter Saur, 2012.

Norris, Dorothy May. *A History of Cataloging and Cataloging Methods.* London: Grafton, 1939.

Noy, Natalya F., and Deborah L. McGuinness. *Ontology Development 101: A Guide to Creating Your First Ontology.* Knowledge Systems Laboratory, Stanford University, 2000. http://www.ksl.stanford.edu/people/dlm/papers/ontology-tutorial-noy-mcguinness-abstract.html.

OCLC. *Bibliographic Formats and Standards.* 4th ed. Dublin, OH: OCLC, 2008. http://www.oclc.org/bibformats/.

OCLC. "Dewey Services." http://www.oclc.org/en/dewey.html.

OCLC. *Online Catalogs: What Users and Librarians Want: An OCLC Report*. Dublin, OH: OCLC, 2009.

O'Connor, Brian C., Jodi Kearns, and Richard L. Anderson. *Doing Things with Information: Beyond Indexing and Abstracting*. Westport, CT: Libraries Unlimited, 2008.

Oh, Kyong Eun. "The Process of Organizing Personal Information." PhD diss., Rutgers University, 2013.

Olson, Hope A. *The Power to Name: Locating the Limits of Subject Representation in Libraries*. Dordrecht, The Netherlands: Kluwer Academic, 2002.

Olson, Hope A., and John J. Boll. *Subject Analysis in Online Catalogs.* 2nd ed. Englewood, CO: Libraries Unlimited, 2001.

Osborn, Andrew D. "The Crisis in Cataloging." In *Foundations of Cataloging: A Sourcebook,* edited by Michael Carpenter and Elaine Svenonius, 90–103. Littleton, CO: Libraries Unlimited, 1985. Originally published in *Library Quarterly* 11, no. 4 (October 1941): 393–411.

Pace, Andrew. “21st Century Library Systems.” *Journal of Library Administration* 49, no. 6 (2009): 641–50.

Patton, Glenn E. “An Introduction to Functional Requirements for Authority Data (FRAD)” and “Understanding the Relationship between FRBR and FRAD.” In *Understanding FRBR: What It Is and How It Will Affect Our Retrieval Tools,* edited by Arlene G. Taylor, 21–33. Westport, CT: Libraries Unlimited, 2007.

Petroski, Henry. *The Book on the Bookshelf.* New York: Knopf, 1999.

Pettee, Julia. “The Subject Approach to Books and the Development of the Dictionary Catalog.” In *Theory of Subject Analysis: A Sourcebook,* edited by Lois Mai Chan, Phyllis A. Richmond, and Elaine Svenonius, 94–98. Littleton, CO: Libraries Unlimited, 1985.

Pitti, Daniel V. “Creator Description: Encoded Archival Context.” *Cataloging & Classification Quarterly* 39, no. 1/2 (2004): 201–26; also published in *Authority Control in Organizing and Accessing Information: Definition and International Experience,* edited by Arlene G. Taylor and Barbara B. Tillett, 201–26. New York: Haworth Information, 2004.

Pomerantz, Jeffrey. *Metadata*. Cambridge, MA: MIT Press, 2015.

Pustejovsky, James, and Amber Stubbs. *Natural Language Annotation for Machine Learning*. Sebastopol, CA: O’Reilly Media, 2013.

Quintarelli, Emanuele. “Folksonomies: Power to the People.” Paper presented at the ISKO Italy-UniMIB Meeting, 2005. http://www.iskoi.org/doc/folksonomies.htm.

Ranganathan, S. R. *Prolegomena to Library Classification.* 3rd ed. London: Asia Publishing House, 1967.

Ranz, Jim. *The Printed Book Catalogue in American Libraries, 1723–1900.* Chicago: American Library Association, 1964.

RDA: Resource Description & Access, developed in a collaborative process led by the Joint Steering Committee for Development of RDA. Chicago: American Library Association, 2010.

Reibel, Daniel B. *Registration Methods for the Small Museum*. Lanham, MD: AltaMira, 2008.

Reynolds, Dennis. *Library Automation: Issues and Applications.* New York: Bowker, 1985.

Riley, Jenn. *Understanding Metadata.* Bethesda, MD: National Information Standards Organization, 2017. http://www.niso.org/apps/group_public/download.php/17446/Understanding%20Metadata.pdf.

Riva, Pat, Patrick Le Boeuf, and Maja Žumer. “IFLA Library Reference Model: A Conceptual Model for Bibliographic Information.” The Hague, Netherlands: IFLA, August 2017. https://www.ifla.org/publications/node/11412.

Robins, David. “Information Architecture, Organizations, and Records Management.” *Records and Information Management Report* 17, no. 3 (March 2001): 1–14.

Roe, Kathleen D. *Arranging & Describing Archives & Manuscripts.* Chicago: Society of American Archivists, 2005.

Rosenfeld, Louis, Peter Morville, and Jorge Arango. *Information Architecture: For the Web and Beyond.* 4th ed. Sebastopol, CA: O’Reilly, 2015.

Rowley, Jennifer, and Richard Hartley. *Organizing Knowledge: An Introduction to Managing Access to Information.* 4th ed. Burlington, VT: Ashgate, 2008.

Russell, Beth M. “Hidden Wisdom and Unseen Treasure: Revisiting Cataloging in Medieval Libraries.” *Cataloging & Classification Quarterly* 26, no. 3 (1998): 21–30.

Saffady, William. *Managing Electronic Records*. New York: Neal-Schuman, 2009.

Salo, Dorothea. "The Humble Index." *The Book of Trogool*. Blog posting, August 25, 2009. http://scientopia.org/blogs/bookoftrogool/2009/08/25/the-humble-index/.

Sauperl, Alenka. *Subject Determination during the Cataloging Process.* Lanham, MD: Scarecrow, 2002.

Schulze, Anna Noakes. "User-Centered Design for Information Professionals." *Journal of Education for Library and Information Science* 42, no. 2 (Spring 2001): 116–22.

Shatford, Sarah. "Analyzing the Subject of a Picture: A Theoretical Approach." *Cataloging & Classification Quarterly* 6, no. 3 (1986): 39–62.

Shepherd, Elizabeth, and Geoffrey Yeo. *Managing Records: A Handbook of Principles and Practice.* London: Facet, 2003.

Smalley, Joseph. "The French Cataloging Code of 1791: A Translation." *Library Quarterly* 61, no. 1 (January 1991): 1–14.

Smallwood, Robert F. *Managing Electronic Records: Methods, Best Practices, and Technologies*. Hoboken, NJ: John Wiley & Sons, 2013.

Smiraglia, Richard P. "Authority Control and the Extent of Derivative Bibliographic Relationships." PhD diss., University of Chicago, 1992.

Smiraglia, Richard P. "Bibliographic Families and Superworks." In *Understanding FRBR: What It Is and How It Will Affect Our Retrieval Tools,* edited by Arlene G. Taylor, 73–86. Westport, CT: Libraries Unlimited, 2007.

Smiraglia, Richard P., ed. *Metadata: A Cataloger's Primer.* New York: Haworth Information, 2005.

Smiraglia, Richard P. *The Nature of "A Work": Implications for the Organization of Knowledge.* Lanham, MD: Scarecrow, 2001.

Snow, Karen. *A Practical Guide to Library of Congress Classification*. Lanham, MD: Rowman & Littlefield, 2017.

Snowden, Dave. "Complex Acts of Knowing: Paradox and Descriptive Self-Awareness." *Bulletin of the American Society for Information Science and Technology* 29, no. 4 (April/May 2003): 23–28.

Soergel, Dagobert. "The Rise of Ontologies or the Reinvention of Classification." *Journal of the American Society for Information Science* 50, no. 12 (October 1999): 1119–20.

Steckel, Mike. "Ranganathan for IAs: An Introduction to the Thought of S. R. Ranganathan for Information Architects." *Boxes and Arrows* (October 7, 2002). http://boxesandarrows.com/ranganathan-for-ias/.

Stoll, Clifford. *Silicon Snake Oil: Second Thoughts on the Information Highway.* New York: Doubleday, 1995.

Strout, Ruth French. "The Development of the Catalog and Cataloging Codes." *Library Quarterly* 26, no. 4 (October 1956): 254–75.

Svenonius, Elaine. *The Intellectual Foundation of Information Organization.* Cambridge, MA: MIT Press, 2000.

Taylor, Arlene G. "Cataloguing." In *World Encyclopedia of Library and Information Services,* 3rd ed., edited by Robert Wedgeworth, 177–81. Chicago: American Library Association, 1993.

Taylor, Arlene G. "The Information Universe: Will We Have Chaos or Control?" *American Libraries* 25, no. 7 (July/August 1994): 629–32.

Taylor, Arlene G. *Introduction to Cataloging and Classification.* 10th ed., with the assistance of David P. Miller. Westport, CT: Libraries Unlimited, 2006.

Taylor, Arlene G. "An Introduction to Functional Requirements for Biblio-graphic Records (FRBR)." In *Understanding FRBR: What It Is and How It Will Affect Our Retrieval Tools,* edited by Arlene G. Taylor, 1–19. Westport, CT: Libraries Unlimited, 2007.

Taylor, Arlene G. "On the Subject of Subjects." *Journal of Academic Librarianship* 21, no. 6 (November 1995): 484–91.

Taylor, Arlene G., ed. *Understanding FRBR: What It Is and How It Will Affect Our Retrieval Tools.* Westport, CT: Libraries Unlimited, 2007.

Taylor, Arlene G., and Daniel N. Joudrey. "Cataloging." In *Encyclopedia of Library and Information Sciences,* 3rd ed., edited by Marcia J. Bates and Mary Niles Maack. New York: Taylor & Francis, 2009.

Taylor, Arlene G., and Barbara Tillett, eds. *Authority Control in Organizing and Accessing Information: Definition and International Experience.* New York: Haworth Information, 2004.

Text Encoding Initiative. "TEI: P5 Guidelines." Version 3.1.0. Last updated on December 15, 2016. http://www.tei-c.org/release/doc/tei-p5-doc/en/html/index.html.

Tillett, Barbara B. "Authority Control: State of the Art and New Perspectives." *Cataloging & Classification Quarterly* 38, no. 3/4 (2004): 23–42; also published in *Authority Control in Organizing and Accessing Information: Definition and International Experience,* edited by Arlene G. Taylor and Barbara B. Tillett, 23–42. New York: Haworth Information, 2004.

Tillett, Barbara B. "Bibliographic Relationships." In *Relationships in the Organization of Knowledge,* edited by Carole A. Bean and Rebecca Green, 9–35. Dordrecht, The Netherlands: Kluwer Academic, 2001.

Tillett, Barbara B. "A Taxonomy of Bibliographic Relationships." *Library Resources & Technical Services* 35, no. 2 (April 1991): 150–58.

UDC Consortium. "About Universal Decimal Classification and the UDC Consortium." http://www.udcc.org/about.htm.

Van Hooland, Seth, and Ruben Verborgh. *Linked Data for Libraries, Archives and Museums: How to Clean, Link and Publish Your Metadata*. Chicago: American Library Association, 2014.

Vellucci, Sherry L. "Metadata and Authority Control." *Library Resources & Technical Services* 44, no. 1 (January 2000): 33–43.

"VRA Core: A Data Standard for the Description of Images and Works of Art and Culture." Library of Congress. https://www.loc.gov/standards/vracore/.

W3C. "Extensible Markup Language." http://www.w3.org/XML/.

W3C. "HTML." http://www.w3.org/html/.

W3C. "HTML5 Tutorial." http://www.w3schools.com/html/.

W3C. "XML Tutorial." http://www.w3schools.com/xml/.

Wallace, Danny P. *Knowledge Management: Historical and Cross-disciplinary Themes.* Westport, CT: Libraries Unlimited, 2007.

Wang, Yongming, and Trevor A. Dawes. "The Next Generation Integrated Library System: A Promise Fulfilled." *Information Technology and Libraries* 31, no. 3 (2012): 76–84.

Webber, Desiree, and Andrew Peters. *Integrated Library Systems: Planning, Selecting, and Implementing*. Santa Barbara, CA: Libraries Unlimited, 2010.

Weinberg, Bella Hass. *Can You Recommend a Good Book on Indexing? Collected Reviews on the Organization of Information.* Medford, NJ: Information Today, 1998.

Weinberg, Bella Hass. "Exhaustivity of Indexes: Books, Journals, and Electronic Full Text." *Key Words* 7, no. 5 (September/October 1999): 1, 6–19.

Weinberger, David. *Everything Is Miscellaneous: The Power of the New Digital Disorder*. New York: Times Books, 2007.

Weiss, Andrew. *Using Massive Digital Libraries: A LITA Guide*. Chicago: American Library Association, 2014.

Wellish, Hans H. "Aboutness and Selection of Topics." *Key Words* 4, no. 2 (March/April 1996): 7–9.

Whittaker, Steve. "Personal Information Management: From Information Consumption to Curation." *Annual Review of Information Science and Technology*, Volume 41. Medford, NJ: Information Today, 2007.

Wilson, Patrick. "The Catalog as Access Mechanism: Background and Concepts." In *Foundations of Cataloging: A Sourcebook,* edited by Michael Carpenter and Elaine Svenonius, 253–68. Littleton, CO: Libraries Unlimited, 1985.

Wilson, Patrick. *Two Kinds of Power: An Essay on Bibliographical Control.* Berkeley: University of California Press, 1968.

Witty, Francis J. "The Beginnings of Indexing and Abstracting: Some Notes Towards a History of Indexing and Abstracting in Antiquity and the Middle Ages." *The Indexer* 8 (1973): 193–98.

Wythe, Deborah, ed. *Museum Archives: An Introduction*. Chicago: Society of American Archivists, 2004.

Yee, Martha M. "Guidelines for OPAC Displays: 1999." In *From Catalog to Gateway: Charting a Course for Future Access: Briefings from the ALCTS Catalog Form and Function Committee,* edited by Bill Sleeman and Pamela Bluh, 83–90. Chicago: Association for Collections & Technical Services, American Library Association, 2005. http://www.pitt.edu/~agtaylor/ala/papers/YeeOPACGuidelines.pdf.

Yee, Martha M., and Sara Shatford Layne. *Improving Online Public Access Catalogs.* Chicago: American Library Association, 1998.

Zeng, Marcia Lei, and Jian Quin. *Metadata*. 2nd ed. Chicago: ALA Neal-Schuman, 2016.

INDEX

Note: Page numbers followed by 'f' or 't' refer to figures and tables.

ABOUT THE AUTHORS

DANIEL N. JOUDREY, MLIS, PhD, is associate professor, School of Library and Information Science, Simmons College, Boston, Massachusetts, where he teaches information organization, subject cataloging and classification, and descriptive cataloging. His published works include *Introduction to Cataloging and Classification*, 11th ed. (with Dr. Arlene G. Taylor and David P. Miller, Libraries Unlimited, 2015); *The Organization of Information*, 3rd edition (with Dr. Taylor, Libraries Unlimited, 2009); "Cataloging," in *The Encyclopedia of Library and Information Science*; and various articles, including some on cataloging education, published in *Cataloging & Classification Quarterly*. With Janis Young, he is co-instructor of the Library of Congress's LCSH Online Training videos. His research interests include aboutness determination, subject access to information, and cataloging education. He holds an MLIS and a PhD from the School of Information Sciences at the University of Pittsburgh, where his studies, guided by Dr. Taylor, focused on subject cataloging, particularly that of determining aboutness.

ARLENE G. TAYLOR, MSLS, PhD, is professor emerita, School of Information Sciences, University of Pittsburgh. Her experience as a library school educator covered more than 30 years, and she was active in the American Library Association for more than 40 years. She is author of 17 books, including *Introduction to Cataloging and Classification*, 11th ed. (with Dr. Daniel N. Joudrey and David P. Miller, Libraries Unlimited, 2015); *The Organization of Information*, 3rd ed. (with Dr. Joudrey, Libraries Unlimited, 2009); and *Understanding FRBR: What It Is and How It Will Affect Our Retrieval Tools* (Libraries Unlimited, 2007). She also has authored more than 70 articles and book chapters. Dr. Taylor received the 1996 Margaret Mann Citation from ALA/ALCTS, the 2000 ALA/Highsmith Library Literature Award, and the 2011 Distinguished Alumnus Award from the Graduate School of Library and Information Science of the University of Illinois. She has given more than 90 guest presentations for international, national, state, and regional library groups.

KATHERINE M. WISSER, MA, MSLS, PhD, is associate professor, School of Library and Information Science, Simmons College, Boston, Massachusetts, where she teaches information organization, archival access and use, metadata, indexing and thesaurus construction, and the history of libraries. Her research interests include metadata, archival description, classification, and the history of libraries. Her publications include articles in *American Archivist*, *Journal of Contemporary Archival Studies*, *Journal of Library Metadata*, and *Journal of Archival Organization*. She served as chair of the working group that established EAC-CPF for the Society of American Archivists.