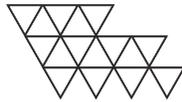


FRBR

BEFORE AND AFTER

A Look at Our Bibliographic Models



KAREN COYLE



An imprint of the American Library Association
Chicago 2016

www.alastore.ala.org

KAREN COYLE is a librarian with over 30 years' experience with library technology, who serves as consultant on a variety of issues relating to digital libraries. She has published dozens of articles and reports, many of which are available at kcoyle.net. She has served on several standards committees, including the MARC standards group (MARBI) and the NISO committee AX for the OpenURL standard, and was an ALA representative on the e-book standards development team that contributed to the ePub standard. She writes and speaks on a wide range of policy areas, including intellectual property, privacy, and public access to information. Her January 2010 issue of *Library Technology Reports*, "Understanding the Semantic Web: Bibliographic Data and Metadata," was awarded the 2011 ALCTS Outstanding Publication Award.

© 2016 by Karen Coyle

Extensive effort has gone into ensuring the reliability of the information in this book; however, the publisher makes no warranty, express or implied, with respect to the material contained herein.

ISBNs

978-0-8389-1345-1 (paper)

978-0-8389-1364-2 (PDF)

978-0-8389-1365-9 (ePub)

978-0-8389-1366-6 (Kindle)

Library of Congress Cataloging-in-Publication Data

Coyle, Karen, author.

FRBR, before and after : a look at our bibliographic models / Karen Coyle.

pages cm

Includes bibliographical references and index.

ISBN 978-0-8389-1345-1 (print : alk. paper) — ISBN 978-0-8389-1365-9 (epub) —

ISBN 978-0-8389-1364-2 (pdf) — ISBN 978-0-8389-1366-6 (kindle) 1. FRBR (Conceptual model) I. Title.

Z666.6.C69 2016

025.3'2—dc23

2015020100

Cover design by Kimberly Thornton. Image © Shutterstock, Inc.

Design composition by Alejandra Diaz in the ITC Galliard Std, Gotham, and Miso typefaces.

© This paper meets the requirements of ANSI/NISO Z39.48-1992 (Permanence of Paper).

Printed in the United States of America

20 19 18 17 16 5 4 3 2 1

CONTENTS

List of Figures vii
Acknowledgments ix
Introduction xi

PART I WORK, MODEL, TECHNOLOGY

ONE The Work / **3**
TWO The Model / **29**
THREE The Technology / **47**

PART II FRBR AND OTHER SOLUTIONS

FOUR FRBR in Context / **65**
FIVE FRBR: Standard for International Sharing / **69**
SIX The Entity-Relation Model / **75**
SEVEN What Is Modeled in FRBR? / **81**
EIGHT Does FRBR Meet FRBR's Objectives? / **97**
NINE Some Issues That Arise in FRBR? / **125**
TEN Bibliographic Description and the Semantic Web / **137**

Afterword 157
Bibliography 161
Index 173

INTRODUCTION

Go to your bookshelf and pull off a book; any book. It may be one you have read many times, or it could be one that is still on your “to read” list. Take a look at it. It may be bound with the flimsy cardboard of a paperback sporting a slick, shiny cover. Or the pages could be held between the cloth-covered boards of a quality hardback. It is unlikely, however, unless you are either very wealthy or very lucky, that your fingers will be touching a fine leather binding.

It is probable that you did not purchase the book for its physical appearance, as appealing as that may be, but for its content. That’s where things get complicated in our story: complicated because it is very hard to say what the content consists of. Words, yes, but you didn’t buy just a set of words, unless your book is a dictionary. No, you bought this book for the story it tells or for the information it imparts. You may have been seeking entertainment, or to learn something new (and happy is the person who gets both!). Although the story or the information came to you as words, you may not be able to recite even a small passage verbatim. We read the words but we remember the meaning, another concept that is difficult to define.

If I ask you some questions about the book, some will be easy to answer, some more difficult. I could ask you for the title, and most likely you know that. The same for the author. You could surely tell me what the book is about, either with a topic (“it’s a history of the Venetian Republic”) or a story (“it’s about a girl who lives on the prairie and what she and her family go through to survive”). Chances are, though, that if I ask you who published the book, you’ll be taking a sneak peek at the title page or the spine to find that information. The place and date of publication will not only be less imprinted in your consciousness but they may actually be a bit hard to find. The precise number of pages is another undeniable fact about the book that may not be on the tip of your tongue.

As a reader, it is the reading experience and what it leaves behind in your memory that makes up the inherent value of the book. And we do know that

readers value their books. There would be no other reason to use the bulk of the wall space of one's home for shelves for book storage, or, when moving to a new home, to pack, lug, and unpack untold pounds of what appears to be inert tree pulp.

Now let's leave books behind and look at other media. Just as many of us love our books, we also have among us many music lovers; people with towering racks of CDs or digital devices chock full of tunes. Here, though, we find some differences from our book story. Ask a music lover the "author" (composer) of a tune and you may be asking the obvious ("Beethoven's 5th symphony") or not ("Santa Claus Is Coming to Town"), even though both pieces of music are easily recognizable when heard. They are recognizable also because, unlike books, we listen to the same piece of music many times, and in different versions. This is a function of the fact that music is performed. Some performances are faithful interpretations of the music, and others, like jazz or digital sampling, are creative distortions of the original.

Music lovers with sufficient talent can reproduce a version of the music either by humming, singing, or playing the music on an instrument. We remember the notes of music in a way that we do not remember the words of a book. But if asked what the music is "about" we are in some difficulty in most cases. Unless the music has a specific story attached to it, such as Sergei Prokofiev's *Peter and the Wolf*, or the teen drama of "Dead Man's Curve," much music does not have a plot or a message that translates to "aboutness."

Other information that only dedicated aficionados of a music genre can relate about their listening choices are date of recording; names of all performers; date of composition; number and types of instruments. Asked what type of music we like, the answers are broad categories like rock, jazz, classical, or country; or sometimes a more specific category, still covering a wide swath: heavy metal; mostly Mozart; Irish folk music; Reggae.

Books and music are two common creative forms that many of us encounter in our everyday lives, and yet what we know about them and how we interact with them are quite different. Now let's look at another creative form: computer games. A player will know the name of the game, the general plot of the play (capture castle, defeat enemy, solve puzzle), and the names of characters. She will also know what capabilities she has as a player (running, jumping, opening doors). If it is a multi-player game, she will know the names of other players—that is, the names they are using in the game. She may not, however, be able to respond to the question "who wrote or created the game?" Games often do not have measurable durations although some have ending points, so asking "how long is it?" may not make sense.

With a movie, on the other hand, the running time for the film is a key element and moviegoers, unless they walk out in the middle of the film, will experience that actual duration. Movies have directors and producers, screenwriters, and hundreds of other participants from makeup artists to caterers. Some directors are famous, but what makes movies *The Movies!* are the stars: the people that you see on the screen. Having seen the film, most people will be able to relate the story and the names of the primary actors. Very few will remember the producer, although his name will have appeared briefly in big letters at the beginning of the film, and even fewer will have noted the screenwriter's name. The name of the studio that produced the film, analogous to the publisher of a book, is rarely noticed.

All of these above-mentioned creative forms are ones to which its users or participants have a certain emotional attachment. There are other kinds of created resources that we seek out but that are less enticing. I'm thinking of items like census figures, standards documents, technical reports, or court proceedings. If asked about authors of these materials, few people outside of librarianship would name courts or government as authors, although they might see them as responsible bodies of some kind. Users of these materials, however, may be keenly aware of the version of the material; a 1950 census is obviously not the same as a year 2000 census, and a version 0.7 of a standards document would be expected to differ from the 2.0 version. Having the latest version may be essential for some functions, although comparisons of figures across time make use of different versions of the data. Knowing that the copy that you have is authentic and has not been altered is another consideration for these materials. For, like census or economic data, a key factor is whether it is formatted for possible number-crunching.



The point of this brief walk through the various resource types is this: given how different these resources are, and how different our relationship to them is, making any general statement about the structure or data elements needed to describe all resources for all users of a library catalog is going to be difficult, if not impossible. And yet, that is exactly what we do on a routine basis: we create records that treat all resource types the same, and for only one definition of “user.” We also ignore or downplay many of the characteristics that are important for users. We often place the names of film actors, when we provide them at all, in a note field that is barely searchable. We also give technical information about data sets and computer files in a note. We give book readers a place of publication and a number of pages but don't give them a clue to the story that the book holds. (“Mentally ill—Fiction” is a subject heading on *Moby Dick*.)

All of this is to point out how varied is our bibliographic universe, and this is without having looked at the differences among users: from novices to experts, children and adults, beach readers and researchers.

Quite clearly, in terms of bibliographic services, one size cannot possibly fit all.

This illustrates the difficulty we have in defining the fundamental nature of the bibliographic “thing,” often called a “work.” And it also illustrates that the users are an element in that definition. It provides an argument for a flexible treatment that can accommodate a range of user approaches and needs, perhaps a modular structure that can be modified to place emphasis on different information for different materials and different users. Why shouldn’t a search on an author return information about the author, including the author’s works? Where was the author born, when did she live, what is she known for? In library catalogs, there is no differentiation between Edgar Allen Poe and Barbara Cartland. This isn’t neutrality, it’s a lack of information. If an item is retrieved on title, there is clearly more that could be said about it than where and when that particular exemplar was published. We present a copy of Charles Darwin’s *On the Origin of Species* with a publication date of 2003 without any further explanation, neither of the importance of the work, nor its own true origins. *On the Origin of Species* is meaningful only if you know what scientific thinking was before Darwin’s discovery, and that this book is the beginning text for the entire science of evolution.

All of this is possible, but only if we can make some fundamental changes in our approach to bibliographic description. A new approach presupposes a redefining of bibliographic description from a fixed, immovable block of data to a set of interrelated information units that can be viewed from different vantage points.

The challenge for us lies in transforming what we can of our data into inter-related “things” without overindulging that metaphor. There are indeed things of interest to be defined for cultural heritage and creative objects, but our universe of operation lacks the precision of, for example, financial data, where every point of information is precisely known, or the calculation of tensile strength in the engineering task of bridge building. What we describe is not easily subject to quantitative testing, and the difference between success and failure is hard to measure. We are fortunate that errors in library catalogs rarely result in death of the user, but we are hindered by a lack of knowledge of our effect on learning and culture. In spite of the attempts in the 1960s to convince the world that one could add the word *science* to *library* and gain a modicum of status, describing information resources remains an art.

We do have some cold, hard facts in our data storehouse, but we also have some squishy bits—some areas where we simply cannot achieve the level of precision

enjoyed by science and engineering. Part of the reason for our imprecision is the durability of our inventory. Unlike a warehouse of electronic gadgets, we don't discard last year's product when the latest offerings arrive. Some of us even keep the old, the ragged, and the unused materials. Our material lacks uniformity: we have books without authors, articles with citations to prior works that no longer exist, artworks without titles, and boxes of papers that we have not yet had time to open much less cogently catalog. There are works with authors whose real identity is hidden behind the mask of a pseudonym or a coy phrase like "Kind Gentlelady of Upper Norwich" as a way to evade censorship or skirt social norms, and thus to confound library users. We have parts of things that should be whole: scattered issues of a journal, volume two of a three-volume publication, the left side of a triptych.

Sometimes to be precise about what we have, we should be equally precise about what we do not, yet we may not know what we do not have. Some number of works are permanently lost due to war, conflagration, neglect, and low budgets. Creative works arise in a cultural and social context, and only an omniscient cataloger could place all of the items owned by the library in their proper place in the extended history of human thought. Omniscient catalogers are, however, in short supply.

Because we cannot achieve omniscience, we have to take advantage of the technologies available to us. At the same time, we need to retain a healthy skepticism against any promises that technologies, on their own, will solve all of the problems of connecting today's seekers to the wealth of recorded intelligence (and sometimes lack thereof) that may be available through a library.

This book looks at the ways that we define the things of the bibliographic world, and in particular how our bibliographic models reflect our technology and the assumed goals of libraries. There is, of course, a history behind this, as well as a present and a future. The first part of the book begins by looking at the concept of the "work" in library cataloging theory, and how that concept has evolved since the mid-nineteenth century to date. Next it talks about models and technology, two areas that need to be understood before taking a long look at where we are today. It then examines the new bibliographic model called Functional Requirements for Bibliographic Records (FRBR) and the technical and social goals that the FRBR Study Group was tasked to address. The FRBR entities are analyzed in some detail. Finally, FRBR as an entity-relation model is compared to a small set of Semantic Web vocabularies that can be seen as variants of the multi-entity bibliographic model that FRBR introduced.



ONE

THE WORK

As librarians became increasingly aware of the concept of the work as a meaningful creative unit separate from the physical package, various members of the profession put forth their ideas on how to define this abstract concept. The best source of information on this aspect of librarianship is Richard Smiraglia's 2001 book, *The Nature of "A Work": Implications for the Organization of Knowledge*.

You might think that a key concept like "work" would be well-understood in libraries, and uncontroversial. You might also assume that libraries would have integrated this basic concept into their services and procedures. Instead, the integration of the work into library practices is, in this second decade of the twenty-first century, still in our future. As Smiraglia has concluded, "a catalog inventory of books must give way to an encyclopedic catalog of works. In this there is no dissent" (Smiraglia 2012).

I suspect that some dissent could always be found within the cataloging community, but it is true that the question of the work had planted itself fully within the

cataloging theory of the mid- to late twentieth century, with Seymour Lubetzky and Patrick Wilson as the most influential theorists of that view.

CREATORS, WORKS, TOPICS

The bibliographic world has its own trinity, which consists of creators, their works, and the place of the works on some conceptual map. None of these concepts is simple, but they vary in their level of complexity. The easiest, from a bibliographic organization point of view, is creators: when neither deceptive nor anonymous, these can often be identified. Next in level of difficulty is the concept of “a work” which is nearly indefinable, yet most of us are quite comfortable with a practical everyday usage of the term. The most complex and difficult concept is that of the topics or subjects of a resource. This latter poses deep philosophical and practical issues, and we have made little change in our approach to subject analysis in the last half century, possibly because there isn’t a clear direction for improving this aspect of our work.

I’m going to assume that the treatment of the creator, as well as other sentient beings who have some role in producing intellectual resources, is fairly well under control. The main activity in this area today is the development of broad and interconnected systems that identify the persons and institutions that are responsible for the production of the resources that are created, disseminated, and curated. None of the existing solutions is perfect—neither library name authority data nor the academic systems that allow researchers to create and maintain their own identities—but progress is being made.

Taking a short digression here, it is worth mentioning that the management of personal identity is hardly a new phenomenon, but it has exploded quantitatively with the advent of social media that puts identity management in the hands of the individual. We still have passports and school records and other identities that are not under our control and which in some cases can represent the unwelcome intrusion of social and political powers. The ability for persons to create, manage, and augment their own identities is a revolution that would have been unimaginable to a small-town dweller just decades ago. In a very short while we have gone from “everyone knows everyone else’s business” to “on the Internet no one knows you are a dog.” We’ve also gone from a limited scope of relationships to being able to broadcast our thoughts around the world. Unfortunately, that doesn’t mean that there are millions who want to listen to us, except perhaps the giant yet impersonal surveillance systems that we now know are hoovering up our bits and bytes, if not actually paying attention to what we have to say.

Socially engineered identity abounds in the modern cultural world. Social and political commentary often takes place in online environments where the authors are pseudonymous. Performers of many types often have a separate public identity from their private identity. In the avant-garde music world, especially where money is not the object and there are few legal contracts that bind relationships, individuals may pass through identities as often as they change their hair color.

Other creative areas have a different approach to identity. Commercial authors' identities are a strong part of their bankability. The best example of this was the attempt by J. K. Rowling, author of the Harry Potter series, to write in a different genre for a different audience, pseudonymously. Sales were modest for the book under the pen name Robert Galbraith. When the true identity of Galbraith was revealed, sales of the book leaped to best seller status immediately. No less a thinker than Michel Foucault suggested that the rise of the author in Western society was precipitated by the need to know who to pay for works, as well as who was to be blamed for them.

Academic writers rely heavily on being properly identified as a work's author so that they will be credited with all of the output upon which their careers depend. This unfortunately has been hindered by the practices of publishers and indexing services, which until recently have not interested themselves in establishing identities, but have been content to record author names without concern for disambiguation. The same person can appear on publications or in bibliographic citations as "John H. Smith," "JH SMITH," "Smith, JH," and so on. Libraries do establish identities for persons, but libraries focus on individually published works, like books, and therefore do not fully cover those academic works that appear in journals.



Returning to subject access to resources, the heyday of library interest in subject access solutions is now quite distant, nearly a century or so past. The development of a combined shelving and classification system in the late nineteenth century by Melvil Dewey was possibly the last great invention in the area of subject access. At the very least, it still informs the methods we use today. Dewey was not alone in his interest in organizing the world of letters topically—that century saw the development of various systems, created by great thinkers such as Paul Otlet, who was responsible for the development of the Universal Decimal System, and Charles A. Cutter, whose Expansive Classification became the basis for the system still in use today in the Library of Congress and other large libraries. In

the twentieth century we had S. R. Ranganathan, the Indian mathematician and librarian who promoted the first fully faceted classification system, and also the members of the British Classification Society of the 1960s and 70s in London. Yet in terms of implementation and innovation in subjects, there has been only a slow evolution of the existing systems like the Dewey Decimal Classification, the Library of Congress Classification, and the Universal Decimal Classification. Ranganathan's brilliant Colon Classification seems to have been too complex to find practical adherents. Limited faceting has been implemented in some library systems, but a fully faceted classification was never employed in Western libraries.

The potential revolution in terms of bibliographic models that is the focus of this book has no effect on subject access. No new subject approaches have been suggested along with the new models for bibliographic description. The proposed descriptive models, from FRBR (Functional Requirements for Bibliographic Records) to BIBFRAME to RDA (Resource Description and Access), each contain a small blank spot where subject access of an undefined nature will presumably be attached to the bibliographic record. We can only speculate on the reasons behind this, but it is abundantly clear that the library descriptive cataloging community has a coherence that is not found in the related subject access area. This may be some accident of history, or it could be related to the feasibility of the tasks that the different groups face. Whatever the reason, we find our profession in the midst of an active discussion of descriptive bibliography, with very little attention going to the task of facilitating access by topic.

WORK: THE WORD, THE MEANING

Words are so beautifully and yet frustratingly meaningful, and the word *work* is a key one in our story. The word has many different uses, and some are relatively precise. You work, she works. A work of art. The works of Shakespeare.

Discussions—or arguments—about the meaning of “work” are part of our philosophical history. Notoriously employed by the post-modern literary critics, the conflict of work versus creator has spawned numerous schools of thought. None of this would matter to those of us involved in public services around works except for that element of “public,” meaning anyone and everyone. A small group of scientists in a tightly-defined research area can agree on a specific use of terminology, or even invent new terms to communicate amongst themselves, but anyone who intends to serve a liberally defined “public” cannot limit her communication to a small group of cognoscenti. There is danger in making use of a term that is already in wide circulation and that has well-established meaning(s), and yet it often is not possible to do otherwise. That is the situation with “work.”

Philosophers, linguists, and cultural critics speak frequently about the meaning of words, but cognitive psychologists actually perform tests. Their focus, however, is less on the individual word but on the concept conveyed and understood by one or more terms. One of the theories that has been the subject of tests in cognitive science is that of degrees of belonging. The easiest way to explain this is to give an example. In an experiment recounted in Gregory L. Murphy's *The Big Book of Concepts* (2004), the subjects are given a list of terms and are asked to put them in order based on the degree to which they answer the question "Is this a fruit?" Although the exact ranking varies, the average ranking comes out something like:

1. orange	6. apricot	11. pineapple	16. pomegranate
2. apple	7. plum	12. blueberry	17. date
3. banana	8. grapes	13. lemon	18. coconut
4. peach	9. strawberry	14. watermelon	19. tomato
5. pear	10. grapefruit	15. honeydew	20. olive

The purpose of this experiment is to show that our categories are not binary; the world is not divided up into fruit/not-fruit, but into a concept of "degrees of fruitness." Few of us would argue with the first couple of items as being high on the "fruitness" scale, and some of us would be surprised to see tomato and olive on the list at all, but not surprised at seeing them at the bottom. How we do this in our brains, and what it means is still an open question. Whether it is subject to some discernable logic, such as commonality of attributes—like sweetness for fruits—is also an open question.

Nor does this ability to categorize bend itself predictably to acquired knowledge. In one experiment, users were asked to rank a group of even numbers based on which they considered the "best" even numbers. Numbers 2, 4, and 8 came out ahead of 34 and 106 (Armstrong 1999). That some even numbers are somehow more even than others is obviously false to anyone with even a minimum background in mathematics, yet the wonderful flexibility of the human brain makes this kind of thinking possible, albeit not necessarily predictable.

If this is a difficult problem with fruits and even numbers, it is an even more difficult problem with less precise concepts. No less an intelligence than Ludwig Wittgenstein set out to prove, in his *Philosophical Investigations*, that we cannot really define unambiguously the concept behind the simple word *game*. That pretty much knocks the wind out of the sails of anyone wanting to use words to communicate anything specific.

We do, however, communicate our ideas and desires and orders using words that represent concepts, and generally our communication is correct. Precision is provided by the context, which also allows us to use terms like *that*, *this*, and *there*. George Kingsley Zipf, who was an early researcher into the statistical analysis of natural language text, showed that there are a relatively few multipurpose words that we use frequently, and presumably in a variety of contexts. These he likens to the general-purpose tools that we keep close to us on our workbench: a hammer, a screwdriver, some pliers. (And it is no coincidence that the saying begins “if all you have is a hammer”) These we can use in many ways. Further out on our workbench, and in the statistical curve that he derived from natural language texts, we find the specialist tools; these are the ones that we use only occasionally, when the general purpose tools are not adequate. Essentially, Zipf provided a logical explanation for the linguistic long tail. The word *bird* will be in the high use area, while *passerine* will be in the long tail (Zipf 1949).

The word *work* is a hammer-like tool, using Zipf’s analogy; it has an imprecise but highly utile meaning. Like many common words in English, it is both a noun and a verb, so to begin with we have to make clear that we are only interested in the noun form. Even with that restriction you can “have work” (meaning employment), “do some hard work” (meaning to labor), or “create a work” (produce a result of some kind). My garden can be a “work of art,” as can a Van Gogh painting. My house is near the “public works” offices of my town, and my bookshelf holds the works of many authors. The word *work* is one of those multipurpose words that supports George Kingsley Zipf’s Principle of Least Effort: it is a word with multiple meanings that, however, makes sense in context.

SOME HISTORY

We live today with an abundance of “product”—there are more books than readers who want them, as evidenced by the copious piles on remainder racks at our bookstores. It wasn’t always thus, of course. Before the advent of printing, each copy was unique and there were few of them. Printing brought exact copies, but it also brought editions, as printers throughout Europe produced their own versions of texts. One European intellectual of the 1500s, Conrad Gessner, felt a need to gain some control over this tsunami of works; he set out to create a universal bibliography of all works in print, but not all of the various editions of the works. Gessner’s *Bibliotheca Universalis* was in part a response to what he saw as wasteful duplication among printers, and he hoped that a list of available works would lead them to concentrate on new works rather than reprinting works

already on the market (Serrai and Serrai 2005). Here it can be said that Gessner obviously did not understand the economics of the book trade.

Libraries, some private, some public, also took advantage of the increased printed book production to grow their collections. One such collection was that of the British Museum Library. In the early 1800s, Anthony Panizzi found himself as head of the British Museum Library with the wonderful title “Keeper of the Printed Books.” This means that there was a parallel position for the other kind of books—manuscripts—and therefore it was necessary to state that “printed books” was a distinct department. We can see this as a kind of microcosm of the transition from precious objects to an abundance that required, as it was later called, “bibliographic control.”

Panizzi had some major problems on his hands. The library’s catalog had been long neglected to the extent that the library had no inventory of its holdings and users could not be sure if the library had the book they sought. The library also had many works in multiple editions coming from the very active English presses. Clearly, Gessner’s goal of stemming the tide of multiple printings of the same work had failed.

The library board had allocated funds for the creation of a new catalog, but not enough to create the catalog that Panizzi felt was needed. This led to the famous showdown between Panizzi and the board as Panizzi explained that a mere “finding list” of authors and titles would not be sufficient for the library to serve its users, nor to efficiently continue to build its collection. The cataloging rules devised by Panizzi specified in each case that the edition be noted by the place of publication and the date, as well as a numbered edition if so stated. (Interestingly, the names of the printers—whom today we would call publishers—were only to be included in his catalog if the printer itself had achieved some level of eminence.)

Some forty years later, when Cutter presented his *Rules for a Dictionary Catalog* in 1876, one of his objects was for the catalog “to assist the user in the choice of a book (G) as to its edition (bibliographical).”

During the decades from 1840 to 1870, the time between Panizzi and Cutter, distinguishing different editions of the same work had become the norm in bibliographic control. Cutter did not discuss whether some users might not care precisely which edition they received, although he did provide an example of the user for whom editions would matter: “for the student, who often wants a particular edition and cares no more for another than he would for an entirely different work.” Cutter’s rules, though, still placed an emphasis on places and dates, and not the publishers themselves: “Print publishers’ names, when it is necessary to give them, in italics after the place” (Cutter 1875).

The rules also acknowledged that the same catalog that served the users also served the library's collection development function, in that the recording of editions was also needed "in the library service, to prevent the rejection of works which are not really duplicates." Duplicate, in 1875, meant the same edition, not the same work.

In my research I have not uncovered the tipping point that led library thinkers like Seymour Lubetzky and Eva Verona to take up the question of the work versus the edition. Yet somehow between the latter part of the nineteenth century and the first half of the twentieth century, it appears that the number of different editions in libraries had become burdensome to users. Although it was still essential to distinguish between editions, it also became important to inform the user that a certain group of editions represented the same work. In just a little over one hundred years we had come full swing from presenting users solely with works, then solely with editions, to needing to gather editions back into their work groups.

THE WORK IN BIBLIOGRAPHIC CONTEXT

We've seen that the term *work* covers a number of different concepts. The difficulty that we have is not with the word, however, but with the meaning that we ascribe to it. Eva Verona, who could be regarded as an early twentieth-century philosopher in the area of cataloging, chose to refer to the focus within the cataloging context as the "bibliographic unit" (Verona 1985). That would distinguish the "item in hand" that is being described from the abstract concept that some wish to be called a "work." Indecs, the metadata model developed in the late twentieth century for digital commerce, referred to "stuff" in its basic diagram, which reads: "People make stuff; people make deals about stuff." This is an interesting punt on defining the exchange of value for labor. (One wonders how Karl Marx would have reacted to such a definition.)

The question of defining the work in the context of library catalogs is multi-fold. Its meaning must be functional, that is, it should serve a purpose. Defining that purpose is not a simple matter. It also needs to communicate readily to the broad and heterogeneous population that both creates catalogs and uses those catalogs. Without dwelling overly on the choice of terms, we can look at the desired functionality expressed by thinkers in the library arena.

Lubetzky's Work View

Seymour Lubetzky was arguably the most influential force in cataloging theory in the twentieth century. He began working at the Library of Congress (LC) www.alastore.ala.org

in 1943, and one of his first assignments was to do a study of the descriptive cataloging rules used by LC at the time, the second edition of the A.L.A. Cataloging Rules, published in 1941. Lubetzky's analysis led to a revision of the rules, issued in 1949. By 1955 he was awarded the Margaret Mann Citation for his contributions to cataloging. He continued to study, publish, and teach as a professor at the School of Library Service at the University of California, Los Angeles. Even after retirement in 1975 he spoke at meetings and participated in discussions. He published his last work in 1999. In the year 1998 the library world feted Lubetzky's one-hundredth birthday with a special symposium. Lubetzky was there. He died in 2003 at the age of 104.

Lubetzky's analysis of the principles of cataloging, published in 1969, became the groundwork for all cataloging rules that have followed. This work greatly influenced the revision of the Anglo-American Cataloguing Rules (AACR) in 1978. Although clearly erudite and studious, Lubetzky's approach to the catalog had a large dose of common sense. In particular, he insisted that the cataloging rules be derived from the functions they were to serve. This was not the case with the 1941 ALA rules that he was first asked to study, which resembled, according to Julia Pettee, "an encyclopedia of pedantic distinctions." (Lubetzky 2001, xiv) Some of Lubetzky's ideas would be considered heretical even today. For example, he decried the repetition of the author between the heading and the statement of responsibility. He also criticized the fact that the information on the card was not placed in order of importance, causing users to scan through unwanted information to look for what served them.

There are two threads in Lubetzky's work that came to the fore at the end of the twentieth century when new bibliographic models were proposed. The first is that the content of the book is not represented by a physical description of the book. This seems obvious, but descriptive cataloging does focus on physicality, and sometimes solely on physicality. Lubetzky argued that the physical "is only a medium through which the work of an author, the product of his mind or skill, is present . . . and that, consequently, the material and the work presented by it are not, and should not be treated as one thing" (Lubetzky 2001,). This is the separation of content (the work) and carrier (the physical medium), although the implementation of this in the library catalog remained (and remains) vague. The second thread is that these physical books (or other media) can be editions of the same work. This establishes a relationship between bibliographic items based on their "workness." Unfortunately exactly how one determines workness was neither defined nor explained. As we know from later efforts, this raises a number of awkward questions about where one work ends and another begins, and whether there are degrees of workness.

INDEX

A

AACR (Anglo-American Cataloguing Rules),
11, 22
AACR2 (Anglo-American Cataloguing Rules,
Second Edition), 24
Abbot, Ezra, 49
accompanying relationships, 21
aggregates, 90, 130–135
AI (artificial intelligence), 54, 59
A.L.A. Cataloging Rules, 11, 22
alternative models using FRBR concepts
 FaBiO (FRBR-aligned Bibliographic
 Ontology), 154–156
 FRBRcore, 148–150
 FRBRoo, 150–153
 <indec> event-oriented model, 153–154
 overview, 147–148
American Library Association, 49
Antelman, Kristin, 119
Apostles of Culture (Garrison), 44
art and artwork cataloging, 122–123
Avram, Henriette, 50, 72

B

Baca, Murtha, 123
Basic Level National Bibliographic Record
 (BLNB), 104–106
basic level of categorization, 17–19
Berners-Lee, Tim, 54
BIBFRAME, 6, 85, 145–147
bibliographic context, works in, 10–20
bibliographic description and the Semantic
 Web

 alternative models using FRBR concepts,
 147–156
 entity relationship modeling and Semantic
 Web models, 138–142
 FRBR in RDF, 142–147
 overview, 137–138
bibliographic families and works, 13–15, 16,
 87, 110
bibliographic function of library catalog,
 39–40
bibliographic relationships, 20–22, 92–93
Bibliotheca Universalis (Gessner), 8
The Big Book of Concepts (Murphy), 7
Bobst Library (New York University), 16
books *versus* works, 12
British Museum Library, 9
Burke Theological Library, 16

C

Calhoun, Karen, 66
card catalogs
 as data models, 30
 out-of-date, 51
 overview, 49–50
cardinality, 58, 59
Carlyle, Allyson, 98
cataloging
 efficiency, 113–117
 rules, Lubetzky's analysis of, 11
 works in cataloging practice, 22–27
Cataloging Cultural Objects (CCO), 122–123
Choi, Yunseon, 126

Clarke, Sherman, 123

classes, 61–62

classification systems, 6. *See also specific classification systems*

cognitive view of works, 17–19

collective titles, 26

collocation, 23, 112–113

Colon Classification, 6

“Composing in Real Time: Jazz Performances as ‘Works’ in the FRBR Model” (Schmidt), 120

computerized catalogs, 50–53

concept, object, event, and place
and CCO, 123
overview, 83, 91–92

conceptual model
FRBR as, 83–86, 157–158
overview, 33

Conceptual Reference Model of the
International Council of Museums
(CIDOC CRM), 150

corporate bodies. *See* persons, corporate bodies, families

Coyle’s cognitive view of works, 17–19

creators, 4–5

Cutter, Charles Ammi, 5, 9, 35–38, 39, 42, 45, 48, 51, 102

D

Data Analysis: The Key to Data Base Design (Perkinson), 73

data models
history of, 30–34
library data models, 34–46
overview, 29–30, 81

data sharing
overview, 117–118
on Semantic Web, 141–142

database management systems, 31, 51–53

database models in RDA, 140–141

Davis, Ian, 148–149

degrees of belonging, 7

Delsey, Tom, 72, 79

derivative relationships, 21

descriptive relationships, 21

development and review of FRBR, 69–74

Dewey, Melvil, 5, 49, 50

Dewey Decimal Classification, 6, 49

digital versions of resources, 116–117

disjoints, 127–128, 129–130, 143

Doerr, Martin, 152

Domanovszky, Ákos, 22

Dublin Core (DC), 60–61

Dublin Core Metadata Community, 144

Dublin Core Metadata Initiative (DCMI), 60

Dublin Core standard for metadata, 60–61, 144

Dulabahn, Elizabeth, 72

Dunsire, Gordon, 82

E

economic reasons for FRBR, 71–72

editions *versus* works, 8–10

endeavor class, 149

entity-relationship (E-R) model
overview, 33–34, 66–67, 75–80
and Semantic Web models, 138–142
of work, expression, manifestation, and item (WEMI), 94–95

equivalence, 58

equivalence relationships, 21

Expansive Classification, 5

expression-dominant model for works, 19–20, 110, 155

expressions
overview, 88–89
users seeking items at level of, 110–111

F

FaBiO (FRBR-aligned Bibliographic Ontology), 154–156

families. *See* persons, corporate bodies, families

film cataloging, 121–122

flat file technology, 30–31

format neutrality, 118–124

Foucault, Michel, 5

FRAD (Functional Requirements for Authority Data), 91, 101

- FRBR: A Guide for the Perplexed* (Maxwell), 130
- FRBR Final Report, 66, 71–73, 77, 84, 85–90, 93, 99–101, 104–105, 106–108, 124, 129–130, 131–132, 157–158
- FRBR (Functional Requirements for Bibliographic Records). *See also* alternative models using FRBR concepts; objectives for FRBR as conceptual model, 83–86 overview, 6, 66–68
- FRBR in RDF
BIBFRAME, 145–147
FRBRer, 142–143
RDA in RDF, 144–145
- “FRBR in the Real World” (Weiss and Shadle), 119
- FRBR Review Group, 82, 142
- FRBR Study Group, 54, 65, 66–67, 70–73, 76–78, 82, 84, 85, 99, 107, 124, 132
- FRBRcore, 148–150
- FRBRer, 142–143
- FRBRoo, 150–153
- “The FRBR Model as Applied to Continuing Resources” (Jones), 120
- FRSAD (Functional Requirements for Subject Authority Data), 101
- functional requirements, 99–104, 157–158
- The Functions and Objects of Author and Title Cataloguing* (Domanovszky), 22
- The Future of the Descriptive Cataloging Rules* (Gorman), 98
- G**
- Galbraith, Robert, 5
- GARE (Guidelines for Authority and Reference Entries), 105
- Garrison, Dee, 44
- Genette, Gerard, 131
- Georgetown University Library, 16
- Gessner, Conrad, 8–9
- GoodReads, 145
- Goodrum, Abby A., 122
- Gorman, Michael, 98
- graphs, 55–56
- Group 1. *See* work, expression, manifestation, and item
- Group 2. *See* persons, corporate bodies, families
- Group 3. *See* concept, object, event, and place
- groupings of entities, 82–83
- H**
- Hendler, James, 54
- hierarchy and inheritance, 125–128
- I**
- identity management, 4–5
- IFLA (International Federation of Library Associations and Institutions), 19, 69
- ILS (integrated library system), 41
- <indec> event-oriented model, 10, 153–154
- inferencing, 58
- inheritance and hierarchy, 125–128
- International Cataloguing Principles, 22, 42, 44, 69, 103
- International Conference on Cataloguing Principles (Paris Principles), 36–38, 45
- International Congress on National Bibliographic Services, 70
- International Society for Music Information Retrieval, 121
- Internet Movie Database (IMDB), 145
- Internet search engines, 40–41
- “Interpreting the Second Objective of the Catalog” (Wilson), 15, 28
- ISBD (International Standards for Bibliographic Description), 69, 70–71, 104–105
- ISBN (International Standard Book Number), 102
- items, 90
- J**
- Joint Steering Committee (JSC), 140–141, 144
- Jones, Ed, 120

K

- Kalf, Ruth, 124
- Krajewski, Marcus, 30, 49
- Krier, Laura, 120

L

- Lassila, Ora, 54
- Lavoie, Brian, 114, 115
- Le Boeuf, Patrick, 106, 152
- Library Bureau, 49
- library catalog
 - bibliographic function of, 39–40
 - within entire library workflow, 41–42
 - goals of, 35–38, 42–45
 - overview, 34–35
 - and users, 42–45
- library catalog technology
 - card catalogs, 49–50
 - computerized catalogs, 50–53
 - database management systems, 51–53
 - historical overview, 48–53
 - MARC records, 50
 - online catalogs, 50–53
 - present and future of, 53–62
 - printed book catalogs, 48–49
- library data models, 34–46
- Library of Congress, 45, 50
- Library of Congress Classification, 6
- library workflow, library catalog within entire, 41–42
- LibraryThing, 112, 145
- linked data, 57–58, 62
- Linked Open Vocabularies project, 150
- logical models, 33
- Lubetzky, Seymour, 4, 10–13, 18, 19, 22, 25, 28, 45, 70, 99
- Lubetzky's work view, 10–13

M

- Madison, Olivia, 70, 73, 74, 107
- manifestations
 - aggregates as, 132
 - attributes, 100–103
 - overview, 89–90

- many-to-many relationships, 76, 128, 133
- map cataloging, 124
- MARC (Machine Readable Cataloging), 50, 51, 53, 78–80
- MARC 21 standard for metadata, 61, 106, 146
- The MARC Pilot Project* (Avram), 50
- MARXML, 53
- Maxwell, Robert, 92, 130
- McEathron, Scott R., 124
- “Mental Models of the Bibliographic Universe” (Pisanski and Žumer), 109
- metadata
 - Dublin Core standard, 60–61, 144
 - MARC 21 standard, 61, 106, 146
- “Modeling Our Understanding, Understanding Our Models” (Renear and Choi), 126
- moving images cataloging, 121–122
- Mumford, L. Quincy, 50
- Murphy, Gregory L., 7
- music cataloging, 120–121
- music uniform titles, 26–27
- MusicBrainz, 145

N

- national libraries, basic bibliographic record for, 104–106
- The Nature of “A Work”: Implications for the Organization of Knowledge* (Smiraglia), 3, 15
- Newman, Richard, 148–149
- NISO (National Information Standards Organization), 42
- Non-Unique Name Assumption (NUNA), 57, 59, 60
- normalization, 33

O

- object-oriented (OO) models, 34, 54, 150–153
- objectives for FRBR
 - cataloging efficiency, 113–117
 - data sharing, 117–118

format neutrality, 118–124
 functional requirements, 99–104
 national level bibliographic record, 104–106
 overview, 97–98
 user needs, 106–113
 OCLC WorldCat database, 16, 28, 45, 112, 114, 115, 116
 Ohio College Library Center, 51
On the Record (Library of Congress), 66
 one-to-many relationships, 76
 O’Neil, Ed, 132
 online catalogs, 50–53
 ontology, 58
 OPAC (Online Public Access Catalogues), 38
 Open World Assumption, 57, 59, 60, 140
 OpenLibrary, 112
 original work, 110
 Otlet, Paul, 5
 out-of-date card catalogs, 51
 OWL (Web Ontology Language), 58–60, 142, 143

P

Panizzi, Anthony, 9, 37, 38, 65
 Panizzi’s rules, 37, 38, 65
Paper Machines (Krajewsky), 49
 paratexts, 131–132
Paratexts: Thresholds of Interpretation (Genette), 131
 Paris Principles, 36–38, 45
 Pearl, Nancy, 40
 Perkinson, Richard C., 73
 persons, corporate bodies, families
 and CCO, 123
 overview, 83, 91
 Pettee, Julia, 11
Philosophical Investigations (Wittgenstein), 7
 physical models, 33
 Pisanski, Jan, 109, 113
 Principle of Least Effort, 7
Principles of Cataloging, Phase I (Lubetzky), 12
 printed book catalogs, 48–49

R

Ranganathan, S. R., 6, 13
 RDA Implementation Database Scenarios, 79
 RDA (Resource Description and Access)
 database models, 140–141
 FRBR implementation in, 66
 overview, 6, 144
 in RDF, 144–145
 standards, development of, 65–68
 Tillett’s analysis of relationships and, 22
 RDF (Resource Description Framework), 53–54, 55–56, 137, 142–147
 reader’s services, 40
 relational database management systems
 (RDBMS), 31–33, 78, 139–140
 Renear, Allen, 126
 Research Libraries Group, 51
 responsible entity class, 149
 review and development of FRBR, 69–74
 Riley, Jenn, 121
 RLIN system, 51
 Rowling, J. K., 5
 Rules for a Dictionary Catalog (Cutter), 9–10

S

Scherzo project, 111–112
 Schmidt, Raymond, 120
 search engines, 40–41
 Semantic Publishing and Referencing (SPAR)
 suite, 154
 Semantic Web. *See also* bibliographic
 description and the Semantic Web
 advantages of, 140
 classes, 61–62
 data sharing, 141–142
 disadvantages of, 140
 entity-relationship modeling and Semantic
 Web models, 138–142
 inferencing, 58
 linked data, 57–58
 Non-Unique Name Assumption, 57, 60
 Open World Assumption, 57, 60
 overview, 34, 54–56
 semiotic view of works, 15–17

semiotics, 15
 sequential relationships, 21
 serials cataloging, 119–120
 Shadle, Steve, 119
 shared characteristics relationships, 21
 Smiraglia, Richard, 3, 15–17, 18, 22, 28, 93, 114
 Smiraglia's semiotic view of works, 15–17
 standards, development of, 65–68
 Stockholm Seminar on Cataloguing, 69
 sub-classes, 152, 154
 subject access, 5–6
 subject class, 149
 subjects, 4. *See also* concept, object, event, and place
 super-classes, 149–150, 152
 Svenonius, Elaine, 72

T

Taniguchi, Shoichi, 18, 19–20, 27, 110, 155
 Taniguchi's expression-dominant model for works, 19–20, 110, 155
 taxonomies, 127
 technology, library catalog. *See* library catalog technology
 Terms of Reference (Delsey and Avram), 72–73, 75, 92, 98, 108
 Tillett, Barbara, 21–22, 70, 72, 77, 87, 93, 132, 133
 triples, 140
 Tucker, Ben, 72
 Turner, James M., 122
Two Kinds of Power (Wilson), 13, 28, 39–40, 44

U

uniform titles, 22–27, 104
 unique material cataloging, 122–123
 Universal Decimal Classification (UDC), 5, 6
 Universal Modeling Language (UML), 34, 76
 University of California library system, 51
 University of Minnesota Libraries, 43
 Unsworth, John, 43
 URL (Uniform Resource Locator), 55

user needs, 42–45, 106–113
 user task analysis, 99–104, 108
 users seeking items at level of expressions, 110–111

V

Variations Digital Music Library 2 Project, 121
 Vellucci, Sherry, 114
 Verona, Eva, 10
 VTLS, 112

W

Web Ontology Language (OWL), 58–60, 142, 143
 Weiss, Paul, 119
What is FRBR? (Tillett), 87
 whole-part relationships, 21
 Wiesenmüller, Heidrun, 133
 Wilson, Patrick, 4, 13–15, 18, 21, 22, 27, 28, 39, 44, 45, 89, 110
 Wilson's bibliographic families and works, 13–15, 16, 87, 110
 Wittgenstein, Ludwig, 7
 work, expression, manifestation, and item (WEMI)
 aggregates, 130–135
 cataloging efficiency and, 113
 and CCO, 123
 diagrams of, 94–95, 128
 disjoint, entities as, 129–130
 E-R model, 94–95
 inheritance and hierarchy, 125–128
 overview, 83, 86
 and user view of the bibliographic universe, 109–110
 Working Group for Functional Requirements for Bibliographic Records, 158–159
 Working Group on Aggregates, 132–134
 works
 aggregates as, 132–133
 attributes, 102
 in bibliographic context, 10–20
versus books, 12
 in cataloging practice, 22–27

- Coyle's cognitive view, 17–19
defined, 6–8
versus editions, 8–10
Lubetzky's work view, 10–13
overview, 3–4, 87–88
and relationships, 20–22
Smiraglia's semiotic view, 15–17
Taniguchi's expression-dominant model,
19–20, 110, 155
Wilson's bibliographic families and, 13–15,
16, 87, 110
World Wide Web, 54–57, 60
World Wide Web Consortium, 67, 137
WorldCat database, 16, 28, 45, 112, 114, 115,
116
- X**
XML version of MARC record, 53
- Y**
Yee, Martha, 112, 121–122
- Z**
Zipf, George Kingsley, 8
Žumer, Maja, 109, 113, 132