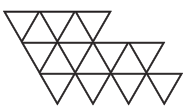
FRBR, Before and After

ALA Editions purchases fund advocacy, awareness, and accreditation programs for library professionals worldwide.

FRBR

BEFORE AND AFTER

A Look at Our Bibliographic Models



KAREN COYLE

ala editions

An imprint of the American Library Association
Chicago 2016

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.

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

FIGURES

ACKNOWLEDGMENTS

I want to thank Patrick Hogan of ALA Publishing for his willingness to experiment with an open access model for this book. A digital version is available at my web site, http://kcoyle.net. I hope, however, that many of you will find the convenience of hard copy worthwhile so that ALA Publishing will be encouraged to do this for other works in the future.

I thank Lynne Cameron for reading through the manuscript and making many highly valuable suggestions. (All remaining errors of fact are, of course, mine.) My conversations and collaboration with Tom Baker have been instrumental to my understanding of FRBR and its relation to RDF. This book began as a series of sketchy and perhaps poorly worded online postings; these brave souls read and commented on that: Kevin M. Randall, Lukas Koster, Heidrun Wiesenmüller, Amanda Cossham, and Joel Kalvesmaki. I hope I did not miss anyone.

Finally, I give thanks to Dewey the Decimal Dog, otherwise known as Dewey 636.7, whose job it was to get me out of my chair and away from the computer at regular intervals. There's nothing like an enthusiastic golden retriever with a tennis ball in his mouth to remind you that there is more to life than FRBR.

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 interrelated "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.

PART I

WORK, MODEL, TECHNOLOGY

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 identify 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 multifold. 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)

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.

Lubetzky did take up the question of books versus works. In his *Principles of Cataloging, Phase I,* issued in 1969 (and never completed), he recognizes that the book itself is a complex entity:

> In summary, then, it must be recognized that, genetically, a book is not an independent entity but represents a particular edition of a particular work by a particular author; and that, consequently, it may be of interest to different users either as a particular edition, or as a representation of a particular work, or as a representation of the work of a particular author. (Lubetzky 2001, 272)

The lack of a definition for works means that some assumptions of the time are not necessarily ones that would be accepted today. Lubetzky was one of the first cataloging theorists to attempt to address the wide range of new media in the cataloging rules, treating non-books as first-class bibliographic entities in their own right, no less worthy of being entered into the catalog than books. In this quote, he allows the concept of "work" to cross the boundaries of physical media, saying "that the same work may be presented in different *media,*" a view that would be greatly qualified today as changes in medium of the type he lists here are considered changes in work.

> Beginning then, with the material cataloged, it is recognized in the revision from the outset that a book, phonorecord, motion picture, or other material is only a medium through which the work of an author, the product of his mind or skill, is present; that the same work may be presented through different media, and in each medium by different editions; and that, consequently, the material and the work presented by it are not, and should not be treated as one thing. (Lubetzky 2001, 199)

Writing in the time of the card catalog, Lubetzky's solutions to the work/edition question are limited to the collocation of works through the use of a "main entry" that consists of the author and the title, or, in the case of editions of a work, the uniform title. Although Lubetzky is considered to have brought the work question to the attention of the library cataloging community, his cataloging rules had little to say about workness, although they did provide significant new approaches to authorship.

In that same 1960 publication, Lubetzky defined a two-part set of primary objectives for the catalog:

> (1) to facilitate the location of a particular publication, and (2) to relate and bring together the editions of a work and the works of an author.

Relating of editions of a work became known as the "second objective," and it was this issue that was addressed by Patrick Wilson not long afterward. The second objective and what it means for the bibliographic model will be covered in a later chapter.

Wilson's Bibliographic Families

Patrick Wilson, professor of Library Science in the University of California at Berkeley School of Library and Information Science, published his book *Two Kinds of Power* in 1968. Although not a focus of the book, he addressed the meaning of the term *work* in the first chapter, "The Bibliographical Universe," in which he defines what he sees as the inhabitants of that universe. It is interesting that by referring to "inhabitants," and not "things," he creates an atmosphere of living beings.

Wilson focuses on texts, and describes the world of letters thus: a person composes a *work,* by ordering letters and words into a *text,* and setting these within an *exemplar.* He makes the point that "these three descriptions are not independent, for he could have produced no work without producing some text, and could have produced no text without producing some permanent or transitory exemplar of the text" (Wilson, 1968, 6). Although they are not independent, each has its own distinct qualities. This may be the first elaboration of the model underlying Group 1 of the Functional Requirements for Bibliographic Records (FRBR), although, as we'll discuss in the modeling section, no two approaches to the inhabitants of the bibliographic biome create exactly the same division of that body.

What Wilson contributes in particular is his own unique definition of the work. He defines a work not as an aspect of a single text, but "a work simply is a group or family of texts." In keeping with the view of beings that inhabit the bibliographic universe, Wilson's works are not static, but the work families develop over time as texts are reproduced or republished in the same or modified form:

> The production of a work is clearly not the writing down of all the members of the family, but is rather the starting of a family, the composing of one or more texts that are the ancestors of later members of the family. (Wilson 1968, 9)

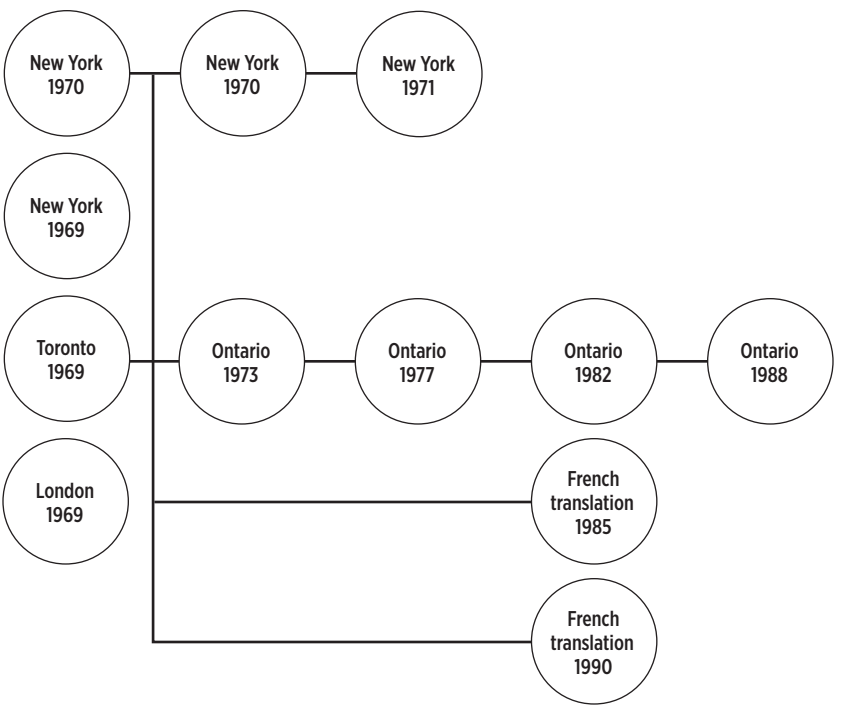
Wilson's view is one possible interpretation of S. R. Ranganathan's statement that "a library is a growing organism." In Wilson's view, the library grows not only in the number of volumes, but with the addition of volumes families grow in a variety of ways. Each addition to the library potentially adds to the familial relationships

that are there, and thus each may alter the nature of the bibliographic family that exists. Works are groups that grow and change over time as new editions or new related works come into being. This of course is a challenge for cataloging because it suggests that catalog entries may not be immutable if relationships are to be included in the catalog. There are relationships from newer resources to older, which could be represented in the description of the newer item only, but the family may grow in different directions. Because items are not necessarily added to the catalog in their order of publication or relation, introduction of new relationships could be disruptive.

In figure 1.1, the "progenitor" is a hardback published in 1969, with a close kin being a paperback in the same year from the same publisher in New York, a Canadian version published in Toronto, and a version published in London. Reprintings of the New York and Toronto versions become children of their respective progenitors. Translations follow, each with the original as "parent" and potentially with children of their own if there are republications of those.

FIGURE 1.1

"The Studhorse Man" as a Wilsonian family, based on Smiraglia 2001



There is no precise definition in Wilson's text to tell us what makes one text a member of a particular family. He considers translations to generally be members of the same family as the progenitor work, but doesn't exclude the possibility that some translations may go so far as to overcome their cultural genealogy and start their own families. It also appears that Wilson did not exclude the idea of a family including adapted works, such as films derived from books. Not being confined by the need to codify his ideas in cataloging rules, he leaves the topic of the work without pinning down a functional definition, and seems to relish the remaining ambiguity: "While there is good reason to distinguish work from text, it is necessary to recognize that the notion of a work is an incorrigibly vague one" (Wilson 1968).

In a 1989 article entitled "Interpreting the Second Objective of the Catalog," Wilson points out something that is obvious once mentioned but often overlooked: that the catalog generally only includes separately published works. Those separate publications often include multiple works, from the prefatory material to the main content, to photographs or illustrations that accompany a text (or to text that accompanies a publication of photographs or illustrations). "By no stretch of the imagination can the author/title catalog be said to give information about all the works available in the library" (Wilson 1989). This of course complicates the study of works, as well as the development of any solutions based on how "works" are defined in the library catalog.

Leaving the work without sharp boundaries is consistent with the remaining theme of his book, in particular his description of the exploitative power, which is individual and contextual and therefore cannot be defined with absolute precision. It is probably his training as a philosopher that allowed him to be comfortable with "incorrigibly vague" concepts; it should come as no surprise that these concepts, then, did not find their way into rules for bibliographical control, where catalogers can't easily sit on the fence over the relationship between a text and a work.

Smiraglia's Semiotic View

Richard Smiraglia has written perhaps the only book on the work question: *The Nature of "A Work": Implications for the Organization of Knowledge* (2001). He covers the various definitions that have arisen in librarianship, more than I include here, but also adds his own, based on the branch of philosophy known as semiotics. Semiotics is a study of meaning, and how meaning is created using signs and symbols. Semiotics is also a study of communication, and therefore touches ever so slightly on the communication theories that have been born out of mathematics and computation. However, the two strike out in very different directions, with semiotics remaining unquantifiable.

Smiraglia calls works "vehicles for communication" and says that "works contain representations of recorded knowledge." Their role is social because they "transport ideas along a human continuum." Works are born as works, both in Smiraglia and Wilson's definitions, yet both allow the "workness" to grow to include new instances as more of the (presumably) same ideas are brought forth as publications.

Smiraglia includes both the ideas and the symbols in his definition of work, whereas Wilson speaks separately of work and text. This speaks to the abstractness of the concept of work; for Smiraglia the work must have been expressed in order to exist. This separation between ideas and expressions is an area where the philosophers of this area diverge.

By taking a semiotic view, Smiraglia includes the reader in his view of the work, and affords the work itself with a cultural and communicative role that changes with each reading (or viewing, or listening). The work is in the eye of the beholder.

> Thus we replace the arbitrariness of the abstract concept of the work with a definitive changeling. Works change over time, they take on new meanings as they are assimilated in cultures, they reflect their perceptions, and they evolve in content and tangibility. (Smiraglia 2001)

Because his view includes communication and culture, his theory can take into account some of the particular characteristics of different kinds of works, such as music, which has the added facet of performance.

Unlike the pure theorists in this summary, Smiraglia conducted quantitative research to discover the extent of work relations in libraries. Using Wilson's concepts of family and progenitor, he sampled the OCLC WorldCat database, New York University's Bobst library, the Georgetown University library, and the Burke Theological Library. Note that these studies were done in 1992 and 1999 and the nature of WorldCat changed considerably after that time, increasing tenfold due to the addition of many millions of bibliographic records from nonmember, and primarily non-US, libraries. The studies were also done on physical libraries, and a combination of physical and digital holdings today could yield different results.

The results in these libraries varied by the type of library: the theological library had numerous older books in its collection, and showed a high rate of "families" in its history area. OCLC, being a union catalog, had the greatest variety of work types. The university libraries each had their specialties, which affected the results of the study. In the end, however, Smiraglia concludes that the "only strong predictor of derivation was the age of the progenitor work" (Smiraglia 2001).

In other words, families develop over time. They also tend to develop more for some genres, like fiction and drama, than for scientific works.

Both Wilson and Smiraglia emphasize that what begins as a new work can give birth to a large family of works through a variety of changes such as revisions, augmentations, performances, and adaptations. Where one draws the line and declares that a new work has been created, however, is not clear.

Coyle's Cognitive View

This is a previously unpublished theory, so I must describe it here at some length. In the section on "Work, the Word," above, I presented a brief explanation of how cognitive science approaches "meaning" and the concepts that are conveyed when we use words to communicate. Cognitive science has studied numerous models of conceptual thinking as part of the human understanding of the world. Concepts have an element of generality/specificity whose exact function in understanding and communication is not yet clear. Regardless of our inability to define how thinking works, every moment provides proof that we do share enough of our conceptual matter to function together in the world. All of this has a strong social component. One of those commonalities is something referred to as the *basic level of categorization,* which means that within a social group we have understood common levels of specificity for things and concepts (Murphy 2004). A simple illustration is this:

Jane and John are walking down the street when they see their neighbor's calico cat. John says: "Hey, there's Fred's cat." Later, at the zoo, Jane says to John: "Take a look at that tiger." Both are felines, yet the words *cat* and *tiger* demonstrate different levels of categorization within our culture, probably based on how common these things are in our shared experience. Each is an understood shorthand for what is obviously a much more complex concept. There is no need to say: "Look, there's a vertebrate mammal of the feline species, sub-species house cat, variety calico, whose owner is Fred," even though that is indeed the case. Instead, "cat" is the level of categorization that allows us to efficiently express a concept that others in our environment will most likely understand. When you type "cat" into the English language Wikipedia, the article that is retrieved represents this same concept of cat as "house cat," while "tiger" gets its own page. This reflects a shared level of categorization in the English-speaking (and Wikipedia-editing) world.

The basic level of categorization is not an absolute, however, but depends on a social context. Experts in a field will have a different basic level than the general

public (e.g., "*Passer domesticus*" and "sparrow") and aficionados amongst themselves will make distinctions that a less interested person will not ("Mercedes-Benz C 215 V6" and"car"). Analogously, librarians will have a shared professional understanding of bibliographic distinctions that is at a more detailed level of categorization than members of the general public.

Lubetzky and others frequently state that a library patron may state that he is looking for a book, when in fact he is interested primarily in the work rather than a specific physical item. The question, though, is what does the patron mean by "book" and what does the librarian mean by "work"? Smiraglia's study of the nature of the work shows that no one single definition of *work* exists among librarians.

If we look at the user view with basic level of categorization in mind, as well as the user's goals, we can then compare that with existing definitions. I'll take as a very simple case a person going to the library to find and check out a book. This person goes to the library and says that he is looking for "the book, *Moby Dick*." Lubetzky and others would say that the user is interested in the work, not a specific physical item. Shoichi Taniguchi (2003) would instead say that the user is interested in the actual text, not the abstraction that is the work. Cognitive science would say that "the book, *Moby Dick*" is a contextual shorthand, most commonly used to refer to a physical (or, today, electronic) book with the text of *Moby Dick*. The user doesn't distinguish between, in Wilson's terms, the work and the text and the exemplar, unless necessary to convey a specific query. The user may not include in her conceptual level that there are variations like translations, annotated editions or works about *Moby Dick* if those are not of interest to her, or not relevant to her immediate context.

The expert user view, for example, that of a professor of American literature who is doing a particular study of technical language in Melville's text, could be very different. Although the level of categorization will be different from that of the casual reader, the focus is still likely to be on the text in a concrete form (on the page or in a digital format). This user may qualify his request as being for "an authoritative version of *Moby Dick*" and may want to check the *bona fides* of the publisher or digitizer. This person is interested at the level of the manifestation, but is still hoping to exit with a real-world object that he can study.

If I say that I have read *Moby Dick*, I am speaking of an experience with a physical book or device that contained the words of Melville and the story those words express. As semioticians might claim, the ideas left in my head from that experience were developed through my experience with the physical book, the text on those pages, what was going on around me during the time that I was reading, and how I interpreted the meaning of that text in my personal context. Nevertheless, a real-world object was encountered.

In both speaking and thinking, we use single and simple terms to represent complex topics, otherwise we could not communicate efficiently. The shorthand used can be fairly imprecise and still support communication. "Have you read *Harry Potter*?" can mean any or all of the books in that "arc" or series. I could answer simply "Yes," meaning that I have read at least one of the works or perhaps all of them. In daily conversation, these shorthands do not cause us problems, in part because we can clarify in the conversation, "All of them?" "Which ones?" But we can also go straight to "What did you think? Good?"

In the cognitive sense, these are not abstractions, but are shared concepts for concrete things that we express with a commonly understood level of categorization that is not too broad to communicate to the other person, but not more specific than it needs to be. The work is often defined as an abstraction, an idea, yet when I ask "Have you read *Harry Potter*?" my question implies inclusion: that the shorthand "*Harry Potter*" represents the whole, and that I am asking my listener about one or more books that the person may have held and read.

In this cognitive model, there is no one definition for "work." It will have meaning within a context and that meaning will often be shared, but not always. The basic level of categorization within that context will vary depending on who is participating in the communication. Librarians are free to develop an expert meaning for the term, but cannot expect that meaning to be shared perfectly with the others. Interaction between libraries and library users of all levels of expertise and knowledge has to mimic the flexibility that humans use unconsciously when communicating, and cannot be so fragile that it is defeated by some degree of ambiguity. For this reason, we should focus on needs and functions, and not on a particular term.

Taniguchi's Expression-Dominant Model

Shoichi Taniguchi is a professor of library science in Japan. He began looking at the models for descriptive cataloging in the mid-1990s at the same time that work was being done by IFLA on the Functional Requirements for Bibliographic Records. Where Lubetzky's general feeling was that most users entering a library looking for a "book" want the "work"; Taniguchi's proposed model placed emphasis on the expressed text, rather than the more abstract work. In fact, Taniguchi's model is probably a better description of the basic level of categorization for texts. He proposes a model of bibliographic description that does not place the work nor the manifestation in the dominant position. He originally called his view "text-dominant," but that was before FRBR's *expression* was defined. His current work is a direct response to FRBR.

In Taniguchi's view, each bibliographic item must be described with a dominant expression entity. Titles, statements of responsibility (including added entries), and edition statements describe the expression; the manifestation bibliographic level contains only those attributes related to publication, physical format, and publication date. His assumption is that the majority of user tasks should be satisfied by the expression in most cases, but could include the work in those situations where work information is normally provided. Using Taniguchi's approach, it is not necessary for a library to create separate entries in its catalog that describe individual manifestations if those are not required by the particular user community. Therefore, time is saved by entering into the catalog the information about the expressions held by the library, and allowing most users to select between manifestations (if more than one exists) at the shelf. The catalog record therefore informs users which expressions the library owns, which is the minimum information needed to fulfill the user tasks. Those few users interested in the details of the manifestation can go on to that level of detail in display.

Taniguchi concludes that very little information about works is included in bibliographic records, although data derived from the manifestation, such as creators, titles, and subjects, is about the work, not the manifestation. The emphasis in cataloging rules is on describing the physical item "in hand," and therefore the dominant entity is the manifestation. He de-emphasizes the physical aspects and organizes his model around the content that the user encounters in the expression of the work.

Although Taniguchi's approach seems at odds with the cataloging of books in libraries, it is easier to appreciate when looking at federated search systems that combine both physical and digital versions of the same materials. This is especially true for journal databases where the particulars of the manifestation have little weight and the types of augmented editions (with added prefaces, illustrations, or commentary) that exist in monograph publishing are virtually unknown. The definition of work may not hold true over all possible bibliographic materials, and may evolve over time as new means of communication develop.

WORKS AND RELATIONSHIPS

Inherent in, but not necessarily explicit in, the definition of works is that bibliographic resources have relationships between them. One of these relationships is "this is a copy or version of the same work," but beyond the question of an exact copy the range and complexity of relationships grows. Most such relationships were not formalized in library catalogs. Instead, for some key relationships, like

translations, supplements, and editions, the coincidence of collocation of entries under the same or similar headings (author, title) was enough to create a logical proximity between related resources. Where headings do not provide collocation, notes are sometimes added.

The new emphasis on works and work relationships spurred discussion of the types of relationships to be found between bibliographic resources. In her 1987 doctoral dissertation, Barbara Tillett undertook a comprehensive study of bibliographic relationships by studying a large set (over two million records) of MARC records from the Library of Congress database. Within these she studied the notes fields that represented statements of relationship, and categorized them. Tillett derived seven types of relationships: equivalence (the same content), derivative (adaptations), descriptive (reviews), whole-part, accompanying, sequential (series), and the more general shared characteristics relationships.

There is no single relationship in Tillett's categorization that translates to "same work" by any of the above definitions. The "equivalence" relationship is limited to copies, reprints, and other republications of precisely the same content. Derivative works include subsequent editions, translations, and adaptations, such as a rewriting of a book for a new audience. Because Tillett studied individual cataloging records produced by the Library of Congress, the bibliographic units in the relationship would be the cataloged publication. As per Wilson's caveat above about the limitation of the library catalog to separately published works, this study covered only some of the bibliographic items held in the library, because it did not include those literary units that were included in larger publications. Tillett did include whole-part relations in her study, but these had to be extrapolated from the existence of contents notes. Clearly the definition of relationships is related to the definition of the unit of bibliographic description, which will become clearer when we look at FRBR and the relationships defined in that model.

In summarizing the seven types of relationships (with their sub-relationships) Tillett wrote in her dissertation: "The primary categories of the above taxonomy meet the criteria of being mutually exclusive and totally exhaustive" (Tillett 1988). Wilson, not only with a more philosophical bent but in his position as a tenured professor, is much less inclined to make such a bold statement. In contemplating relationships, Wilson notes that he does not believe that there is a finite set of relationships, and thus discourages attempts to define such a set. In a practical application of relationships to bibliographic units, the truth is probably somewhere between these two views, with some set of relationships covering the majority of useful relationships, but always allowing for expansion as more is learned or as the nature of catalogs changes.

In the years since the important work that Tillett did to categorize relationships, the possibility that relationships should become incorporated more thoroughly into cataloging rules has gained traction. Her analysis influenced both the development of FRBR as well as that of the cataloging rules, Resource Description and Access (RDA), both of which Tillett was involved in creating.

WORKS IN CATALOGING PRACTICE

Without actually defining the difference between a "book" and a "work," both terms are used in the International Cataloguing Principles of 1961. The key to their use leads us back to Seymour Lubetzky, who, according to Richard Smiraglia and others, greatly influenced the creation of the 1961 principles. The use of *work* in the International Cataloguing Principles seems quite natural on the surface. The functions of the catalog include both "whether the library contains a particular book specified . . . by its author" as well as "which works by a particular author." The term *work* here presumably has the sense of "oeuvre," in the broad meaning of that concept. The Principles state that "The main entry for works entered under title may be either under the title as printed in the book, with an added entry under a uniform title, or under a uniform title." The uniform title is a contrived title that brings together some members of a bibliographic family. The instructions leave the definition of a work and when it should be represented in the catalog to the discretion of the cataloger, all along avoiding any need to tackle the very difficult task of defining what a work is.

By creating a special work title that would be assigned to all instances of the work in the descriptive cataloging, all editions of the same work would be collocated. In his 1989 article that primarily echoes the thinking in Ákos Domanovszky's 1975 book *The Functions and Objects of Author and Title Cataloguing,* Patrick Wilson suggests that one could go beyond recording merely the same edition of a work, but could form a family of works that could include strongly related texts, such as supplements, commentaries, and continuations. This view begins to approach Wilson's desire that a catalog make explicit the relationships between items in the library, with "same work" as only one possible relationship. The relationship "same work" (which may also extend to "same expression" or "same text") is implemented in library catalogs using the mechanism of the *uniform title*. First introduced in the A.L.A. Cataloging Rules of 1941, the uniform title gained additional prominence in the editions of the Anglo-American Cataloguing Rules (AACR). Unfortunately, the uniform title has been applied very unevenly in libraries, and this is at least in part due to the problem of scope.

The purpose of the uniform title is to collocate, that is, bring together in the same place, the versions of a single work. "Collocation" in library cataloging takes place through the relative position of the items in the alphabetically ordered list of the catalog. To overcome differences in how names of creators and titles of works are presented in actual publications, collocation within the ordered list is accomplished by using standardized "headings." These are controlled text strings for the bibliographic data that will be represented in the catalog, such as the names of authors, titles, and subjects. Collocation may sound simple, but in fact there are numerous adjustments that must be made in order to bring together items that the cataloging rules deem to be the same bibliographically. In particular, the collocation of works requires the cataloger not only to identify that different resources represent the same work, but also to provide a heading that will bring the works together in the catalog.

Collocation for works fails in some cases in spite of the normalization of author names because titles of publications of the same work can vary. In modern works this is most often true for translations:

The magic mountain
La montagne magique
Der Zauberberg

Older and ancient works, such as the works of Shakespeare or early sagas that were written before their language or dialect was normalized, may also have titles that have varied over time, like:

Hamlet
Hamlet, Prince of Denmark
The tragedie of Hamlet, Prince of Denmarke

To collocate these in the catalog as variations of a single work, an additional title is added between the author and the title of the printed book. This is called a "uniform title" and it serves as a normalized title that represents the bibliographic work. Where known, the uniform title represents the title of the original publication of the work. In other cases, the title is a selected title, such as "Hamlet," that contains the commonly known name of a work that was published under many different names, especially in its early period. The uniform title can also contain the language of the translation and/or the date of publication, to distinguish between different versions or editions.

```
Mann, Thomas
 Der Zauberberg
Mann, Thomas
 [Zauberberg. English]
 The magic mountain
Mann, Thomas
 [Zauberberg. French]
 La montagne magique
Shakespeare, William
 Hamlet
Shakespeare, William
 [Hamlet]
 Hamlet, Prince of Denmark
Shakespeare, William
 [Hamlet]
 The tragedy of Hamlet, Prince of Denmark
Shakespeare, William
 [Hamlet. Italian]
 Amleto
```

The uniform title, shown here between square brackets, represents the work with a "work title" combined as needed with something that distinguishes between different versions. In the above case that distinction is made with the language of translation, but for some works that appear in different versions in the same language, such as the works of Shakespeare, the expression may be represented by either a date or both a language and a date.

However, in the current cataloging rules, any publications whose title would be the same as the uniform title are not given a uniform title, and the majority of publications have only a single edition, and thus need no uniform title. AACR2 explains it this way:

> The need to use uniform titles varies from one catalogue to another and varies within one catalogue. Base the decision whether to use uniform titles in a particular instance on:
>
> a) how well the work is known b) how many manifestations of the work are involved c) whether the main entry is under title d) whether the work was originally in another language e) the extent to which the catalogue is used for research purposes.

> Although the rules in this chapter are stated as instructions, apply them according to the policy of the cataloguing agency. (Anglo-American Cataloguing Rules, 1978)

As you can see, the exceptions to the creation of a title for a work are both numerous and subjective. Bringing out the "workness" of a resource is the exception rather than the rule, and many libraries make little or no use of uniform titles for the work.

The first exception is that any item that has been published in only one edition or in only one language is not assigned a work title. Even the main proponent of identifying works, Seymour Lubetzky, stated that "wherever an author is identified in his works by one particular name and a work is represented under one title only" nothing more needs to be done to identify the author and the work.

In addition, the different editions of a work are not given a work title in cases where the titles of the editions do not interfere with collocation, as in reprintings or updated editions:

> Eysenck, Michael W., and Mark T. Keane. *Cognitive Psychology: A Student's Handbook*. Hove [u.a.]: Psychology Press, 2010. 6th edition
> Eysenck, Michael W., and Mark T. Keane. *Cognitive Psychology: A Student's Handbook*. Hove [u.a.]: Psychology Press, 2007. 5th edition
> Eysenck, Michael, and Mark T. Keane. *Cognitive Psychology: A Student's Handbook*. Hove: Psychology Press, 2003. 4th edition

Thus, different editions or versions of a work (or members of a work family) are only identified through a heading in those cases where the work title is needed to collocate the entries. If they already collocate by the coincidence of having the same titles, no work is identified.

Note also that, as shown above in the Thomas Mann example, the catalog entry for the item *Der Zauberberg* does not require a uniform title because the uniform title would be the same as the title of the book. This complicates the rules for sorting in catalogs because it requires a cascading sort of uneven membership, where the "real" title must sort before the uniform title that contains the exact same characters.

The uniform title is a great illustration of the tension between serving the individual library's users and the efficiency that can be gained through massive sharing of cataloging copy. Although allowing each library to make its own decisions as to when to bring out the "workness" of a resource is sensible both

from a question of workflow and user service, it has a definite effect on data sharing. What makes a work useful or necessary in one library could be a distinct hindrance in another. A library may have copies of Tolstoy's *War and Peace* in English, Spanish, and Chinese, but not in Russian because the library does not serve a Russian-speaking population. Therefore, each translation can be found under the title in the translated language, which is logically where readers would look to find the book:

Guerra y paz War and peace 戰爭與和平

but it may not be useful to also include an entry under the original Russian title, война и мир. Yet making that decision and adjusting the cataloging copy for the individual library has a cost in terms of cataloger time.

The concept of work in library catalog data is currently unevenly applied in practice. Individual libraries or library groups can and do opportunistically decide whether to make use of this feature based on the criteria in the cataloging rules, plus the perceived needs of their users and the capabilities of their catalog software. Key to the upcoming sections on FRBR is the fact that prior to FRBR, the work and the expression were considered bibliographically significant only under certain circumstances. In part this was because the creation of a specific heading for the work had effects on the catalog and the user experience that were either deemed unnecessary or even detrimental to the users of that library.

So far I've spoken of only one kind of work or "uniform" title. There are two others. The first is the collective title, like "Complete works" or "Selections." The second is the particular type of uniform title used for music materials. Both of these perform the same collocation function that is the basis of the work title, but they have significantly different meanings. The collective title identifies a particular type of publication, often not used on the title of the piece. My own observation is that this is unevenly used, even in large libraries. The music title, however, is a thing unto itself, and is probably the most successful application of work titles to a bibliographic group. Music is in many ways a special case because, unlike texts, musical compositions often do not have a single distinctive title. In addition, we experience music through performances, not through the original creation of the composer. And, last but not least, recorded music is most often packaged by publishers with two or more musical pieces per package, meaning

that there is extensive use of the "added entry," an author/title heading that essentially has a part/whole relationship to the main bibliographic entity.

Music uniform titles are crafted descriptions of the music piece, which sometimes ignore what most of us would consider a true title for the piece. For example, Beethoven's symphony known as the "Eroica" (but also as Beethoven's Third symphony) is given this uniform title:

Symphonies, no. 3, op. 55, E flat major.

No one would consider this artificial construction as the proper name of the symphony. Yet the method neatly orders music—at least classical music—and overcomes the lack of uniform practice in naming such works: "The Eroica," "Beethoven's Symphony #3," "Beethoven's Third Symphony," "Sinfonie in Es-Dur," "Symphonie no 3 en mi bémol majeur," and many more.

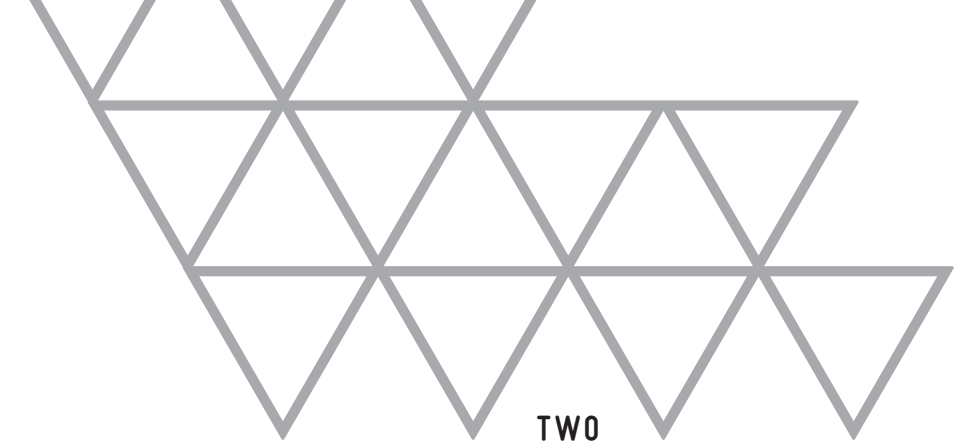
Although important conceptually, as we've seen here, direct presentation of the work in cataloging is limited to a relatively small number of cases in libraries today. Taniguchi points out that in current cataloging the work does not "perform a key role in describing an item being cataloged, although its existence is supposed to be a prerequisite in making a bibliographic description." Catalogers simultaneously describe the item in hand and extrapolate some degree of "workness" in assigning headings, but only when that seems called for. Moving to a bibliographic description that recognizes the work sufficiently to reveal the bibliographic families that Patrick Wilson describes means a significant change in cataloging practice. Recognizing those works in a way that the bibliographic families can be identified and offered to users as such is a much more difficult task plagued with some deep philosophical and practical questions. Among these is that of defining the boundaries within which bibliographic decisions take place. By elevating the bibliographic discourse from publications to works, the universe expands from the physical library and the item in hand to an essentially unbounded abstraction. Exactly where that abstraction should be addressed, whether within the inventory of a single library or in some aggregated bibliographic layer that is not limited to a library's holdings, is a question that has not been answered, and often is not even asked.

Summary

The preceding definitions of the work are not to be taken as exhaustive nor conclusive. These definitions hopefully provide a bit of perspective for when we talk about the more functional approach of current bibliographic models.

There are many issues that are not addressed here but that pertain to how we define works. There is no in-depth discussion of whether all resources have some degree of workness. The studies cited here were either limited to text, or text and music. Recorded knowledge comes in other forms, including aerial photography, topographic maps, and scientific datasets. Whether each of these is also imbued with the work quality as defined by our thinkers is not clear. In his 1989 article "Second Objective of the Catalog," Patrick Wilson identifies some resources that are publications without being works, such as collections of shorter works between a single set of covers. This is not a universally accepted point of view, as we will see in the section on FRBR and aggregates. Although this opens up the possibility that there are "non-works" it does not provide criteria that we could use to divide the works from the non-works.

There is also little discussion of the domain of discourse in these definitions. Patrick Wilson's *Two Kinds of Power* addresses the need to define a domain, but rather oddly defines the domain of the library catalog as being the items in the library as well as items being considered for inclusion in the library. In other areas, he speaks of the "bibliographic universe," which is the broadest view one can take. How the library catalog intersects with the bibliographic universe is not stated, nor is what this means for the definition of the work. Lubetzky and Smiraglia's investigations generally use the context of the library catalog, and in Smiraglia's quantitative studies the boundaries for the work families are always inside a single catalog (even though that single catalog, WorldCat, can be an aggregation of many library catalogs). The question that isn't answered is whether there is a work family if the members of that family are not present in your catalog. Yet how we model our universe depends on having a clear answer to that question.



TWO

THE MODEL

There are various reasons to create models of the real world, mostly having to do with the difficulty of manipulating the real world directly. Architects create models of buildings they have designed, car-makers create clay models of new automobile designs, and chemists create physical models to represent molecules. Oftentimes our model of the world is not a physical model but a symbolic data model. These models are abstractions of the real world, and their resemblance to reality is conceptual rather than physical. Unlike an architect's or car-makers model, a data model doesn't physically resemble the thing we are modeling. This necessary abstraction from the real world makes the development of data models complex and prone to error. There are numerous competing techniques for the development of data models that help guide one in this difficult task. These techniques are used even by modeling experts.

Models generally begin with a macro view of the area of interest, such as growth plans for a city. They place the subject of the model in context and state general goals. The next step is often articulation of use cases. Use cases can be more or

less specific, but they should state in clear terms what functionality the data in the model must support. The use case for a car is that a person can get into it and drive it from one place to another. Because one might drive a car after dark, it has to have lights that one can turn on that illuminate the road well enough for traveling. There must be a steering mechanism so the driver can turn the car in needed directions. Only when this type of functionality is articulated does the design team then get down to the details of implementation. In data models, the macro level is the enterprise. If the enterprise is large and complex, more than one system may be needed to serve all of its needs, and therefore sub-units with distinct boundaries become the area being modeled. The overall goals of the enterprise ("build cars and sell a lot of them") are the context for the model of a data system that serves all or some portion of the enterprise.

SHORT HISTORY OF DATA MODELS

We can credit libraries with developing some of the earliest data models with the development of the card catalog. Card catalogs were indeed "paper machines," as Markus Krajewski (2011) calls them, with interchangeable parts and a predictable retrieval method. The punched card had essentially the same functionality as a manual card file, only it could be run through a machine process that acted on the information encoded on the cards. Punched cards had limited capabilities because they only held eighty (actually seventy-two after eight were dedicated to sequencing) character positions.

The next advance was the ability to store the previously encoded punch-card data inside the computer itself. As computers became more powerful, the limitation of seventy-two characters per line was lifted, and we got an automated spreadsheet that looked not unlike the ledger book of olden days. If you are accustomed to working with spreadsheets, you may be familiar with data that has a form like this:

NAME	STREET	CITY	STATE	ZIP
John Smith	123 Main St.	Anytown	New York	10101
Mary Jones	33 High Road	Sometown	California	93003
Jane Doe	77 Lower Road	Anytown	New York	10101
James Roe	989 Norton Pl	Anytown	New York	10102

Spreadsheets are called "flat file" technology because they are simply a list of entries, one after the other, in a single file. You can search spreadsheets, sort them, and extract selected data from them. However, once the amount of data

becomes very large—as would be needed for banking or to manage a large warehouse—the spreadsheet technology is not efficient enough to produce results in a short enough amount of time to make use of the data of the enterprise "in real time." If you don't want to have to wait overnight to get an answer to your query, you need better technology.

Flat files can become very bulky with repeated data. For example, if you have a list of customers and the products they have purchased, you quickly get a large file where some data is represented many times. If a customer buys more than one product, you need to list the customer again for each product purchased.

NAME	STREET	CITY	STATE	ZIP	PRODUCT	QTY.
John Smith	123 Main St.	Anytown	New York	10101	X12	2
John Smith	123 Main St.	Anytown	New York	10101	X13	1
Mary Jones	33 High Road	Sometown	California	93003	X12	1
Mary Jones	33 High Road	Sometown	California	93003	P38	6

Every repeated element requires an entire new entry in the table. You can see how a file could grow quickly in size. The solution, at least the solution in the last decades of the twentieth century, would be to use a "database management system" rather than a spreadsheet. Early database management systems used a hierarchical model that could query particular paths in order to arrive at results. Like the classified library shelving system, these hierarchies forced designers to provide one and only one place for each information unit, which naturally cut off some possible data combinations at the same time that it facilitated others. In our example above, the model would need either to store customers in a hierarchy under products, or products under customers. Neither would be ideal, and there would still be repetition at the lower levels of the hierarchy. By the 1970s a new type of database management system was developed that was much more flexible than the hierarchical system: it was called a "relational database management system," or RDBMS.

The primary goals of a relational database are to eliminate duplication of the same information in the database, and to create relationships among bits of information such that it would be possible to approach the data from almost any starting point and still retrieve what you need. A relational analysis of the first spreadsheet shown above would begin by noting the duplication in the city, state, and zip code columns. That could then be designed as seen in Figure 2.1.

Each separate entry in a relational database is called a table, and figure 2.1 shows a mock-up of a database design based on the spread sheet, but now with two tables.

FIGURE 2.1

Data redesigned as two database tables

CUSTOMERS		
NAME	STREET	ZIP
JSmith	123 Main	10101
MJones	33 High	93003
JDoe	77 Lower	10101

ZIP+		
ZIP	CITY	STATE
10101	Anytown	New York
93003	Sometown	California

There is still duplication here, within the city, state, zip-code table. The three columns for city, state, and zip code have a built-in relationship: the same zip code is always related to the same city and state, but the same city and state can have multiple zip codes. Therefore, the zip code can be considered a "key" for the city and state, and those can be placed in a separate table.

The purchase information related to customers becomes an additional set of tables that have relationships with the customer information. The logical database design therefore becomes something like in figure 2.2, although actual designs are generally much more complex.

FIGURE 2.2

Data redesigned as three database tables

CUSTOMERS			
ID	NAME	STREET	ZIP
1	JSmith	123 Main	10101
2	MJones	33 High	93003
3	JDoe	77 Lower	10101

PURCHASES		
CUST_ID	PRODUCT	QTY
1	X12	2
1	X13	1
2	X12	1
3	P38	6

ZIP+		
ZIP	CITY	STATE
10101	Anytown	New York
93003	Sometown	California

This process of analysis of the data to eliminate duplication is called "normalization." Normalization is generally the second or third step in a multistep analysis. This analysis might use a technique called "entity-relation modeling." Imagine that you work in a highly complex enterprise that is planning to computerize its operations. You have hundreds of employees in offices that each manage the data for a different function of the enterprise, such as manufacturing, purchasing, sales, and personnel. You wish to integrate all of these so that each office has access to the information it needs, and the data moves through the workflow without being duplicated (or lost). You ideally don't begin by tossing in all of your spreadsheets and paper files and beginning a normalization of your data. Instead, your model begins with a macro view that would make sense to management and nontechnical employees. From that you move into more detail, finally looking at individual data elements and the capacity of the actual database management system that you will employ.

Entity-relation (E-R) modeling is a technique developed in the 1970s and 80s to describe the elements of the data universe that you wish to organize and their relationships to each other. The technique was developed specifically to aid in the design of relational databases, although it has value in other data mapping situations as well. The first step in E-R modeling provides a conceptual view of your data. A *conceptual model* serves to define the data "things" (entities) that your business works with, and how they relate to each other in the bigger picture. Once the conceptual model is well understood, the process moves on to the creation of a *logical model*. This is where you complete the list of data elements, and define what type of data value will be stored for each data element (text, date, currency). This is the phase where you discover duplicate data coming in from different functions and perform normalization on the data. As you can imagine, the resulting picture can be very complex, and may vary considerably from the conceptual model. A *physical model* is the final step in database design, and may be combined with the logical model into a single step. The physical model should reflect the actual database structure and contents.

The "conceptual model" of E-R modeling is not conceptual in the philosophical or cognitive science definition of "conceptual," but is a first step toward development of an actual data processing system. In philosophy or cognitive science, concepts can be imprecise, changeable over time and within different contexts, and probably could not be accurately developed into anything as mechanical as a database management system. In E-R modeling, the concepts define the main categories of things that must be described in the data in order to support the functional requirements of the system, and the relationships between them. Quite often the conceptual model is much simpler than the subsequent logical model.

E-R modeling is still used, as are relational databases, although in the 1990s a new model of data processing was developed, called "object-oriented" (OO). Object-oriented concepts are behind the programming languages C++ and Java, as well as being the basis for current languages like Python and Ruby. Object-oriented design makes extensive use of *classes* to gather data elements and processing routines that are shared by data types. OO classes can function as modular routines that encapsulate existing programming code, thus protecting that part of the code from changes made to the program elsewhere. A new design notation was developed to help developers who were working with OO models: the Universal Modeling Language, or UML. UML can be seen as an evolution of E-R modeling; it is possible to create E-R models using UML, but UML supports over a dozen types of modeling needs, including structure modeling, behavior or process modeling, and interaction modeling. Other than the extensive use of classes, one of the more significant differences between OO and E-R designs is that object-oriented programming and design often focuses on dynamic processes rather than static views of data. OO data is more like a factory than a finished product.

The next leap forward in data-planning and design is that brought on by the development of the Semantic Web. At this writing, the Semantic Web revolution is still in progress, and data designers are just beginning to gain experience with this new way of looking at the data we manage and share. The Semantic Web uses the concept of a web or graph of data, with the Internet as its underlying technology. The Semantic Web emphasizes growth and interconnections between data that can come from different environments. Although it is being used in business applications, the Semantic Web is oriented more toward discovery and knowledge enhancement than control. This will be covered more comprehensively in the chapter on technology.

LIBRARY DATA MODELS

Libraries have a number of functions that are served by their data systems: acquisitions and fund accounting, personnel administration, inventory control, user identification, and, of course, the library catalog. The actual function of the library catalog is where I will focus our discussion of modeling here, but before I do I want to talk about the bigger picture in libraries.

If you grab a book on data modeling, it will give you steps to take that lead from functions performed by employees all of the way to a database design that allows them to do their jobs with the help of automation. These books assume

that the database that is being designed will be built. That seems like an obvious thing to bring up, after all why would you be designing a database unless you intended to build it? However, this is exactly the situation that libraries are in: libraries do not build systems, and they have only minor control over the systems that are built for them. For this reason, what few modeling exercises take place in libraries are quite different from those that we see coming from the enterprise information technology sector.

There is one aspect of library information management that overshadows all others, at least in library data theory, and that is the catalog of the library's holdings. To some extent, the catalog *is* the library, because it is itself a model, in metadata, of the essence of the library: the information it offers. The library catalog is to the library as the architect's miniature is to the real building. You would think, then, that there would be a large body of work around the model of the catalog and its implementation in technology. That is not the case, however. There is a body of work on the theory and practice of cataloging, but it is distinctly separate from any discussion of satisfying those goals in technology design. The library profession models its data, but not the system solution that uses that data. This leads to an awkward situation where the goals of cataloging may not be the same as the functions of the catalog as implemented.

Goals of the Catalog

In 1875, Charles Ammi Cutter stated the goals of the library catalog as:

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)

Cutter defines a catalog as a "list of books which is arranged on some definite plan." He distinguishes the catalog from a bibliography in that a catalog is a "list of books in some library or collection," while a bibliography is a list of books around some other organizing principle, such as subject, place or period. To Cutter, the catalog's main goal is to be "an efficient instrument."

Cutter's list of goals could be considered a high-level set of use cases. What is not articulated here, but obviously was clear enough to him that he could develop his cataloging rules, was exactly how the catalog is to provide this functionality. There is nothing here to say how users will find an author, or what it means that the catalog will "show what the library has." Of course, Cutter was working nearly one hundred years before the concept of systems analysis was common among modelers, so to point out this shortcoming is not a criticism of the great man, but does show how modeling has changed as a concept.

In 1961, the International Conference on Cataloguing Principles (known as the "Paris Principles") gave these as the functions of the catalog:

> The catalogue 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 the 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.

The similarities between these functions and Cutter's goals are striking. The 1961 Paris Principles, written ninety years after Cutter, change his wording somewhat but have essentially the same meaning: the purpose of the catalog is to provide an identity for the resources in the library by a small set of known qualities, such as the author of the work, or the title, that a catalog user can employ to discover if the library has a copy of the item sought. There is no question that these principles adhere to the distinction between bibliography and a library catalog that was defined by Cutter. The library catalog is a sophisticated finding aid. Unspoken but implicit is that users can also discover what a library does not have because it will not appear in the catalog.

Significantly, the Paris Principles do not mention subject or genre access, both of which were included in Cutter's requirements for the catalog. Cutter's rules

devoted fifteen pages to describing subject access, less than ten percent of the total, although Cutter conceded the exact subject description methodology to sources external to his cataloging rules. The scope of the Paris Principles was limited to entries by authors' names and titles (and the latter only when author entry was for some reason not available). In this sense, the Paris Principles can be seen as an updated version of Panizzi's rules, which preceded them by over a century. Both require author entry where the author name is available, define title entry for those works without authors, and deal with the form of the author's name and a set of exceptions. And no more. These principles comprise only a portion what one generally considers a complete catalog for users.

The most recent version of these principles was issued in 2009, nearly 50 years after the original Paris Principles and over 125 years since Cutter laid out his goals.

> 4. Objectives and Functions of the Catalogue
>
> The catalogue should be an effective and efficient instrument that enables a user:
>
> 4.1 to find bibliographic resource in a collection as the result of a search using attributes or relationships of the resources:
> 4.1.1. to **find** a single resource
> 4.1.2. to **find** sets of resources representing
> all resources belonging to the same work
> all resources embodying the same expression
> all resources exemplifying the same manifestation
> all resources associated with a given person, family, or corporate body
> all resources on a given subject
> all resources defined by other criteria (language, place of publication, publication date, content type carrier type, etc.), usually as a secondary limiting of a search result;
> 4.2. to **identify** a bibliographic resource or agent . . . ;
> 4.3. to **select** a bibliographic resource that is appropriate to the user's needs . . . ;
> 4.4. to **acquire** or **obtain** access to an item described . . . ;
> 4.5. to **navigate** within a catalog and beyond . . .

The change here is significant, and is entirely due to the fact that this version of the Paris Principles follows (temporally and philosophically) the entities described in the Functional Requirements for Bibliographic Records (FRBR). The "book" has been replaced with the FRBR bibliographic entities "work, expression, manifestation," even though those are not defined anywhere in this version of the

document. Subjects return in this edition, although as we will see they are actually given short shrift in the FRBR model. The principles also include an interesting smattering of "additional access points" that don't appear to have any particular theoretical basis, such as "bibliographic record identifiers," "language of expression," and "content type." None of these are defined or explained, and the suggestion is that these may be used as a "limiting device for a search." Such devices are found in some online catalogs, but there doesn't appear to be a philosophical basis for their existence in the Principles.

Although user-seeking behavior was implied in previous versions (users "found" in Cutter, and "ascertained" in 1961), this 2009 version includes the user tasks defined in FRBR: find, identify, select, and obtain. It also adds the concept of sets, an acknowledgment of what the introduction to that document refers to as the "OPAC (Online Public Access Catalogues)" technology in wide use. The term *set* refers to the technology of retrieval that, based on a query, returns a selected group of entries that meet the criteria of the query. This may seem to be a small change, yet in fact the change from the linear, alphabetic (or "dictionary" catalog, as Cutter would have it) is a change of great import that is hardly acknowledged in the practice of bibliography.

This is undoubtedly not the first time that you will have seen Cutter's rules, because his rules for a dictionary catalog continue to be widely quoted as the basis for library cataloging today. To some this is proof that there are strong, underlying purposes to the catalog that have withstood the test of time. On the other hand, it seems unlikely that Cutter's objects of the catalog are sufficient for today's information seekers.

In 1875, when Cutter's rules were published, a very large library was one that held 500,000 volumes, and most libraries were much smaller. Information seeking in a collection of that size is clearly different from information seeking in a library holding millions of books and tens of thousands of motion pictures and pieces of recorded music, and also provides integrated access to tens or hundreds of millions of indexed articles. The library user of 1875 was of course also significantly different from the library user of the twenty-first century. Some of the arguments launched against Panizzi's plan to create a detailed catalog of books in the British Museum Catalog were that any reasonably educated gentleman came to the library knowing exactly what he sought, and therefore the additional information in the catalog was unnecessary.

In the midst of all of this orthodoxy around library catalog goals, some interesting ideas came from outside of the cataloging community. One particularly unorthodox thinker was Professor Patrick Wilson, and his exposition of a concept he called "two kinds of power."

Patrick Wilson's *Two Kinds of Power,* published in 1968, and introduced in chapter 1, is a book that is often mentioned in library literature but whose message does not seem to have disseminated through library and cataloging thinking. If it had, our catalogs today might have a very different character. A professor of Library Science at the University of California at Berkeley, Wilson's background was in philosophy, and his book took a distinctly philosophical approach to the question he posed, which most likely limited its effect on the practical world of librarianship. Because he approached his argument from all points of view, argued for and against, and did not derive any conclusions that could be implemented, there would need to be a rather long road from Wilson's philosophy to actual cataloging code.

Wilson takes up the question of the goals of what he calls "bibliography," albeit applied to the bibliographical function of the library catalog. The message in the book, as I read it, is fairly straightforward once all of Wilson's points and counterpoints are contemplated. He begins by stating something that seems obvious but is also generally missing from cataloging theory, which is that people read for a purpose, and that they come to the library looking for the best text (Wilson limits his argument to texts) for their purpose. This user need was not included in Cutter's description of the catalog as an "efficient instrument." By Wilson's definition, Cutter (and the international principles that followed) dealt only with one catalog function: "bibliographic control." Wilson suggests that in fact there are two such functions, which he calls "powers": the first is the evaluatively neutral description of books, which was first defined by Cutter and is the role of descriptive cataloging, called "bibliographic control"; the second is the appraisal of texts, which facilitates the exploitation of the texts by the reader. This has traditionally been limited to the realm of scholarly bibliography or of "recommender" services.

This definition pits the library catalog against the tradition of bibliography, the latter being an analysis of the resources on a topic, organized in terms of the potential exploitation of the text: general works, foundational works, or works organized by school of thought. These address what he sees as the user's goal, which is "the ability to make the best use of a body of writings." The second power is, in Wilson's view, the superior capability. He describes descriptive control somewhat sarcastically as "an ability to line up a population

of writings in any arbitrary order, and make the population march to one's command" (Wilson 1968).

If one accepts Wilson's statement that users wish to find the text that best suits their need, it would be hard to argue that libraries should not be trying to present the best texts to users. This, however, goes counter to the stated goal of the library catalog as that of bibliographic control, and when the topic of "best" is broached, one finds an element of neutrality fundamentalism that pervades some library thinking. This is of course irreconcilable with the fact that some of these same institutions pride themselves on their "readers' services" that help readers find exactly the right book for them. The popularity of the readers' advisory books of Nancy Pearl and social networks like Goodreads, where users share their evaluations of texts, show that there is a great interest on the part of library users and other readers to be pointed to "good books." How users or reference librarians are supposed to identify the right books for them in a catalog that treats all resources neutrally is not addressed by cataloging theory.

Wilson's analysis presages the search and retrieval capabilities of Internet search engines like Google, Bing, and Yahoo. He also writes that power of bibliography is greatest if it extends over the entire bibliographic universe, not just a single selection (one universal library as opposed to the local collection); that the user is better served the fewer retrieved items must be reviewed before satisfying the user's request (as in targeted ranking); and that direct access to the text is a greater power than restrictive use (open access).

Due to the philosophical nature of the book, one has to tease out these brilliant ideas; they are not laid out as headlines or clear conclusions. Yet in the text Wilson may have laid out a new direction for libraries decades before those same principles were discovered by Internet entrepreneurs using new technologies. Imagine if Internet search engines had the same goals as library catalogs and designed their products to cater to only those users who came to the search box knowing either the title or the author of the document they were seeking. Not only is that not the goal of these systems, but they do not even assume that the search engine user is even aware that any documents satisfying their need exist. This is the difference between seeing information space as a finite set of items on a shelf, versus an ever-changing, nearly infinite set of unknowns. The setting of boundaries around the library collection is one of the tenets of library cataloging goals—to define exactly what the library does and does not have. Although such an inventory is clearly needed, it is a mistake to also assume that this inventory and its boundaries is what interests today's information seeker. Cutter's goals for the catalog were written at a time when the information world was still

contained within a relatively small number of published texts, and even fewer of those were available to information seekers at any given time and place. Although users may have entered a library seeking information, the only possible way to pose the question at that time was "do you have a book on?" A person facing the nearly blank Google home page is free to ask "is there anything out there about my topic?" without having to predetermine the limitations that may exist in the information resources available on that topic. Failure in these systems is undoubtedly a common occurrence, but the failure in the library catalog comes about by limiting the questions the user can ask, and limiting, by design, the utility of the response.

The Larger Context

I began this section saying that a model begins at a macro level. A model that covers the library catalog and the user interaction with that catalog is clearly already focused on a small slice of both the library's functioning and on the activities of the user. You could argue that this is a self-contained unit that is well-defined, but it is easy to prove otherwise.

Many library management functions revolve around the resources owned or controlled by the library, such as acquisitions and collection development. This is the basis behind the idea of the "integrated library system," or ILS. There is a workflow not unlike that of a business where resources are selected for purchase, added to budgets, paid out as expenses, received as goods, processed, and stored. Prior to the integration of these workflows, separate systems had their own separate databases, and these often carried information duplicating that of other areas of the library's management. The integrated system brought at least some of these data stores together, resulting in less duplication and greater efficiency. Given this, it would seem only sensible that the catalog would be studied within the entire library workflow. If it were, there would be goals like:

- Show what the library has on order.
- Allow the input of minimum records for items under review.
- Keep a record of requested inter-library loans for future purchasing decisions.
- Manage statistics about use and co-use of materials.

The catalog that is described in the cataloging rules and in the models of catalog data does not acknowledge the existence of library management functions.

Not that the library cataloging rules would necessarily be the correct place for information like account management, circulation statistics, or serials receipts, but the failure to place the catalog in the larger context means that there isn't a place in the model for the interaction of these necessarily connected functions.

At the same time, look at any request for proposal for an integrated library system, and neither cataloging goals nor users receive much attention, just as the needs of library systems are not addressed in cataloging rules. This split between the goals of the user catalog and the goals of the library as a place of business is also visible in the standards environment. Technical standards are developed by the National Information Standards Organization (NISO). There are standards for circulation data, for statistics, for automated data retrieval, for recording licenses, for serials management, and a number of identifiers. The base format for recording the catalog data is also a NISO standard, but the specific format used is managed elsewhere, at the Library of Congress. Although NISO has a work area called "Discovery to Delivery" this area does not include any direct interaction with the cataloging rules, which are developed by a separate and independent organization. NISO also does not have standards that would overlap with the library cataloging rules, nor with the goals for the catalog.

The upshot is that libraries have moved into the twenty-first century with nineteenth century user service goals, at least as far as information seeking in the library catalog is involved. Although today's systems could provide a wide variety of user services, there is no interaction between technology standards development and cataloging standards. The addition of "all resources defined by other criteria (language, place of publication, publication date, content type, carrier type, etc.), usually as a secondary limiting of a search result"; to the 2009 International Catalog Principles is in its way proof of how distant cataloging is from technology design. It is ironic that almost none of the "other criteria" that are actually used in systems and that allow limiting by such come from the cataloging rules. In practice, these systems make use of the fields in the machine-readable record standard that the cataloging rules do not describe, much less mandate, as catalog information.. The information is usable in this way precisely because it is coded information designed for use by computers, not as visible information for human users.

The User in the Model

The catalog goals also provide a very narrow view of the user's interaction with the library. We will see this again when we look more closely at FRBR, even though its "find, identify, select, obtain" appears to be broader than Cutter's "find a book of which _____ is known."

First, what do the goals tell us about the user? The first thing is that some users come to the library looking for a known item. This is indisputable. Whether they really know what they are looking for is another question, and we have seen that online systems use technologies like query completion and "did you mean . . . ?" because this is a common problem.

Next we have the user finding sets that represent logical groupings, such as all of the works of a single author. Once again, it appears that users need to come to the library with this information, because nowhere is it stated that the system should offer these sets through some other mechanism. In fact, many systems do, by allowing users to click on a linked heading and retrieve everything associated with that heading, but because there has been no definition of the functions of the catalog, this isn't something we can assume.

What is key about these goals, however, is that they limit themselves to the user finding an entry in the catalog (albeit FRBR goes on to having the user obtain the item represented there). A study done by the University of Minnesota Libraries in 2006 (UMN 2006) took a much broader view of their users and user needs. They asked their faculty and graduate student users questions like "Where do you work when you are conducting research?" "How do you share source materials?" Just these two questions already reveal quite a lot: the librarians are not assuming that one conducts research in the library, and acknowledge that many people work in teams or groups that share resources among themselves. They also asked about library use: how often do these users visit the physical library, and how often do they visit the library web site, and what do they do there?

The authors of the report (who modestly remain anonymous) then developed a model to describe what they had learned. They borrowed the core of their model from a humanities researcher, John Unsworth, who described the primitives of humanities research as *discover, gather, create,* and *share*. Of these, only discover is usually seen as directly related to the library, and many, perhaps even most, discoveries take place outside of the library catalog. Yet if your view is that libraries support the research function, then all of these primitives could possibly have some interaction with the library. The *share* primitive includes teaching, and the library may be directly connected to the course management system such that course materials are shared through library functions. The *gather* function includes acquiring and organizing, which might mean library support of bibliographic tools. And the *create* function could be supported through shared annotation tools, which could be especially important in those disciplines where research is done through collaborative work.

Libraries have recently begun to take a role in the storage and sharing of research data. Oftentimes institutional repositories for the storage and delivery

of research papers written by the faculty of an institution are also managed by the library. In many of these, the library users are not using the library to find materials, but are instead providing resources that the library will manage. Even if those materials do not go through the same cataloging process as more traditional library holdings, it would be hard to argue that they should not be equally available for searching.

Although libraries have taken on many of these functions, and some of them do interact directly with the library catalog, they are not included in the objectives and functions of the catalog listed in the International Cataloguing Principles. Those principles expound an unfortunately narrow view of the catalog, isolated from the user services that modern libraries are endeavoring to provide.

The objectives of the catalog say little about the users themselves and why they would come to the library seeking resources. Wilson addresses this in *Two Kinds of Power* when he states that it is obvious that people are looking for the best book for their needs or desires. I characterize the traditional library catalog goals as beginning with "a man walks up to a catalog. . . ." Nothing before or after the interaction with the catalog is under consideration. What those objectives do is put a tight fence around the freedom of a person to then ask the question that would satisfy their need. Because of how the catalog is designed, the question "Do you have a good book on dogs?" is not going to result in an answer, although it is, in Wilson's view, simply illogical to think that someone would ask the question "Do you have a book on dogs that I will find insufficient for my needs?" It also seems unlikely that someone would come looking for "a list of books on dogs where there isn't enough information for me to determine which meet my needs."

From this view it becomes clear that the objectives of the catalog are not stated in terms of satisfying the user's query, but to delineate what queries can be made, and to manage the expectations for what responses will be experienced. Library instruction in universities teaches users what they can—and cannot—ask of various resources available through the library, precisely because none of them can answer the question: "Do you have what I need?" Bibliographic research is often a tedious and unsatisfying task. Course syllabi and best-seller lists exist precisely because this is so.

The question comes down to the moral role of the library. As historian Dee Garrison pointed out in her book *Apostles of Culture* (1979), in the early twentieth century libraries saw their role as uplifting the ignorant masses by providing them

with "good books." The library as neutral keeper of the "stuff" came about later, but arguments for moral education still come forward around allowing comic books into the library and providing unfiltered access to the Internet. Thus the debate over whether the library provides what the user does want, or provides what the user should want, continues. In the area of the catalog, however, the solution appears to be to provide only discernible facts about resources.

Patrick Wilson later addressed a topic of more specific interest to catalog theory, and that is the identification of the library resources that represent that same "literary unit." Lubetzky referred to this as cataloging's "second objective." Whereas it would be a notable expansion of bibliographic description for libraries to attempt to fulfill Wilson's second kind of power, library catalogs already include some bibliographic relationships between the items in the library and beyond. Both Cutter and the original Paris Principles include the identification of the edition of a book as a basic function of the catalog. This goes beyond the mere description of individual items to adding certain bibliographic relationships between items where appropriate. Unlike Wilson's second kind of power, this idea has actually gained some traction.

In any functional model it is necessary to define a clear scope of operation: what are the boundaries within which this model will operate? Cutter was clear in his objectives that his rules applied to the catalog of a library, and served to show what books the library did hold, and, by deduction, what books it did not. He had a clear universe for his rules, and it was the single library. The challenge to the neat, finite boundaries of single library's walls came about twenty-five years later when the Library of Congress began distributing sets of catalog cards to libraries across the United States. With this seemingly small gesture, the closed walls of the individual library catalog were breached.

Since then libraries have had to seek a balance between the efficiency of bibliographic data sharing and the desire to serve their unique population of users. The development of combined catalogs of the holdings of multiple libraries, including the massive WorldCat database containing the holdings of tens of thousands of libraries, makes the creation of a boundary for a bibliographic data model all the more elusive. Creating a viable model when such a key question is unresolved is difficult if not impossible.

THREE

THE TECHNOLOGY

Today when we say "technology" it is often shorthand for "computer technology." The Technology section of a newspaper reports on Silicon Valley news and reviews the latest consumer gadgets that are powered by bits and bytes. Of course this is not the only technology in our lives, but it is the one that defines our modern age. A century and a half ago, the defining technology was electricity and all things electric. The light bulb was literally the bright idea of the day. Today we have LED light bulbs that we can control with a smartphone app, turning on the lights when we are still on our way home, or creating a romantic atmosphere by changing the color and intensity of the light at the touch of a screen.

If we move back in time we see ages defined by their technological innovations: steam power, water power, or the precision use of metals that made it possible to create accurate timepieces and to automate the production of fine cloth. We can go back to the printing press, clearly a defining technology for all that came after it. Printing technology depended both on innovations with metals and also on

the development of paper-making techniques that greatly improved on previous writing surfaces, like sheepskin, papyrus, wax, clay, and stone.

Basically, it's technology all of the way back—back to fire and the first stone axes. We naturally take for granted the technologies that precede our own age, and we marvel at the ones that are new.

Libraries of course have been technology-based from the beginning of their history. The earliest libraries that we know of were furnished with writings in the form of scrolls. Medieval libraries held bound manuscripts. The big leap forward was the Gutenberg revolution and the concomitant increase in the production of copies of texts. The number of books not only increased but they also become more affordable as a result of their abundance. Other technologies also had effects on libraries, such as the aforementioned development of electric lighting, which reduced the threat of fire and allowed readers to make use of the library outside of daylight hours.

In the eighteenth and nineteenth centuries, not only were more copies of books produced than ever before, but the numbers of new writings and new editions also grew. Library holdings thus increased as well, which led to difficulties in keeping up with an inventory of the items held by the library. Today we assume that every library has a catalog, but even in the 1800s some libraries had no actual record of their holdings or relied on a brief author list. Much "finding" done in libraries at the time relied on the memory of the librarian. Charles Ammi Cutter, writing about the catalog of the Harvard College Library in 1869, took pity on the librarian overseeing a collection of 20,000 books without a proper catalog, who had to attempt to answer subject-based queries using only his own knowledge of the content of the collection.

The library catalog technology of Cutter's day was a printed book. Printed book catalogs had the same advantages as books themselves: they could be produced in multiple copies and were highly portable. A library could give a copy of its catalog to another library, thus making it possible for users to discover, at a distance, that a library had the item sought. The disadvantages of the printed book catalog, however, became more serious as library collections grew and the rate of growth increased. A library catalog needed near-constant updating. Yet the time required to produce a printed book catalog in an era in which printing required that each page be typeset meant that the printed catalog could be seriously out of date as it came off the printing press. Updating such a catalog meant reprinting it in its entirely, or staving off an expensive new edition by producing supplementary volumes of newly acquired works, which then made searching quite tedious.

In the mid-1800s the library card catalog was already winning hearts and minds. Cutter attributed the development of the card catalog to Ezra Abbot, head of the Harvard College Library, in 1861 (Cutter 1869). Although neither the book catalog nor the card catalog meets all needs as efficiently as one would desire, the card catalog had already proven itself as an up-to-date instrument for library users and librarians alike. German professor Markus Krajewsky, in his book on the history of card files, *Paper Machines* (2011), shows that cards on paper slips had been used in earlier times, in particular by the early bibliographers and encyclopedists who needed to create an ordered presentation of a large number of individual entries. It was libraries, however, that demonstrated how useful and flexible the card catalog could be.

Cards were lauded by Melvil Dewey in his introduction to early editions of his Decimal Classification, although his classification and "relativ index" in no way required the use of a card system. However, the "Co-Operation Committee" of the newly formed American Library Association announced its decision on the standardization of the catalog card in *Library Journal* in 1877; not coincidentally, Dewey's library service company, The Library Bureau, founded in 1876, was poised to provide the cards to libraries at a cost lower than custom-produced card stock. The Library Bureau soon branched out into the provision of catalog furniture and a variety of card-based products for a growing business records market. In fact, before long providing cards to libraries was only a small portion of The Library Bureau's revenue as businesses and other enterprises in the United States and Europe turned to card systems for record-keeping. Krajewski considers these card systems the early precursors of the computerized database because of the way that they atomized data into manipulatable units, and also allowed the reordering of the data for different purposes.

It should be obvious that both the book catalog and the card catalog were themselves technologies, each with different affordances. They also were affected by related technological developments, such as changes in printing technologies. The typewriter brought greater uniformity to the card catalog than even the neatest "library hand" could, and undoubtedly increased the amount of information that one could squeeze into the approximate 3" x 5" surface. When the Library of Congress developed printed card sets using the ALA standard size and offered them for sale starting in 1902, the use of the card catalog in US libraries was solidified.

After Melvil Dewey, the person who had the greatest effect on library technology was Henriette Avram, creator of the Machine Readable Cataloging (MARC) format. This was not only an innovation in terms of library technology, it was generally innovative in terms of the computing capability of the time. In the mid-1960s, when MARC was under development, computer capabilities for handling textual data were very crude. To get an idea of what I mean, look at the mailing label on any of your magazines. You will see upper-case characters only, limited field sizes, and often a lack of punctuation beyond perhaps a hash mark for apartment numbers. This is what all data looked like in 1965. However, libraries needed to represent actual document titles and author names, and languages other than English. This meant that the library data record needed to have variable length fields, full punctuation, and diacritical marks. Avram delivered a standard that was definitely ahead of its time.

Although the primary focus of the standard was to automate the printing of cards for the Library of Congress's card service, Avram worked with staff at Library of Congress and other libraries involved in the project to leverage the MARC record for other uses, such as the local printing of "new books" lists. To make these possible the standard included non-text fields (in MARC known as "fixed fields") that could be easily used by simple sort routines. The idea that the catalog could be created as a computerized, online access system from such records was still a decade away, but Librarian of Congress L. Quincy Mumford announced in his foreword to Avram's 1968 document *The MARC Pilot Project* that MARC records would be distributed beginning in that year, and that this "should facilitate the development of automation throughout the entire library community." And it did.

Melvil Dewey did not anticipate the availability of the Library of Congress printed card service when he proposed the standardization of the library catalog card, yet it was precisely that standardization that made it possible for libraries across America to add LC printed cards to their catalogs. Likewise, Henriette Avram did not anticipate the creation of the computerized online catalog during her early work on the MARC format, but it was the existence of years of library cataloging in a machine-readable form that made the OPAC a possibility.

The next development in library catalog technology was the creation of that computerized catalog. It would be great to be able to say that the move from the card catalog to the online catalog was done mainly with the library user's needs

in mind. That wasn't my experience working on the University of California's online catalog in the early 1980s. The primary motivators for that catalog were the need to share information about library holdings across the entire state university system (and the associated cost savings), and to move away from the expense and inefficiency of card production and the maintenance of very large card catalogs. At the time that the library developed the first union catalog, which was generated from less than a half dozen years of MARC records created on the systems provided by the Ohio College Library Center (later known solely as OCLC) and the Research Libraries' Group's RLIN system, the larger libraries in the University of California systems were running from 100,000 to 150,000 cards behind filing into their massive card catalogs. This meant that cards entered the catalog about three months after the book was cataloged and shelved. For a major research library, having a catalog that was three months out of date, and only promising to get worse as library staffing decreased due to budget cuts, made the online catalog solution a necessity.

We, and by "we" I mean all of us in library technology during this time, created those first systems using the data we had, not the data we would have liked to have. The MARC records that we worked with were in essence the by-product of card production. And now, some thirty-five years later, we are still using much the same data even though information technology has changed greatly during that time, potentially affording us many opportunities for innovation. Quite possibly the greatest mistake made in the last two to three decades was failing to create a new data standard that would be more suited to modern technology and less an imitation of the library card in machine-readable form. The MARC record, designed as a format to carry bibliographic data to the printer, was hardly suited to database storage and manipulation. That doesn't mean that databases couldn't be created, and to be sure all online catalogs have made use of database technology of some type to provide search and display capabilities, but it is far from ideal from an information technology standpoint.

The real problem is the mismatch of the models between the carefully groomed text of the catalog entry and the inherent functionality of the database management system. The catalog data was designed to be encountered in an alphabetical sequence of full headings, read as strings from left to right; strings such as "Tolkien, J. R. R. (John Ronald Reuel), 1892–1973" or "Tonkin, Gulf of, Region—Commerce—History—Congresses." Following the catalog model of which Charles Cutter was a primary proponent, the headings for authors, titles, and subjects are designed to be filed together in alphabetical order in a "dictionary catalog."

Database management systems, which are essential to permit efficient searching of large amounts of data, work on an entirely different principle from the sequential file. A database management system is able to perform what is called "random access," which is the ability to go seemingly directly to the entry or entries that match the query. (The actual internal mechanism of this access is quite operationally complex.) These entries are then "retrieved," which means that they are pulled from the database as a set. A set of retrieved entries may be from radically different areas of the alphabetical sequence, and once retrieved are no longer in the context intended by the alphabetical catalog.

Database management systems include the ability to treat each word in a sentence or string as a separate searchable unit. This has been accepted as a positive development by searchers, and is now such a common feature of searching that today most do not realize that it was a novelty to their elders. No longer does a search have to begin at the same left-anchored entry determined by the library cataloging rules; no longer does the user need to know to search "Tonkin, Gulf of . . ." and not "Gulf of Tonkin." Oddly enough, in spite of the overwhelming use of keyword searching in library catalogs, which has been shown to be preferred by users even when a left-anchored string search was also available, library cataloging has continued its focus on headings designed for discovery via an alphabetical sequence. The entire basis of the discovery mechanism addressed by the cataloging rules has been rendered moot in the design of online catalogs, and the basic functioning of the online catalog does not implement the intended model of the card catalog. Parallel to the oft-voiced complaint that systems developers simply did not understand the intention of the catalog, the misunderstanding actually goes both ways: significant difference in retrieval methods, that is, sequential discovery on headings versus set retrieval on keywords, did not lead to any adaptation of cataloging output to facilitate the goals of the catalog in the new computerized environment. Library systems remain at this impasse, some three-and-a-half decades into the history of the online catalog. The reasons for this are complex and have both social and economic components.

It is not easy to explain why change was not made at this point in our technology history, but at least one of the factors was the failure to understand that cataloging is a response to technical possibilities. Whether the catalog is a book, a card file, or an online system, it can only be implemented as an available technology. Unlike most other communities, the library community continues to develop some key data standards that it claims are "technology neutral." It is, however, obvious that any data created today will be processed by computers, will be managed by database software, will be searched using database search

capabilities, and will be accessed by users over a computer network. One ignores this technology at great peril.

THE PRESENT AND FUTURE

We have made the error in the past of moving to new technologies without examining the fit between our data and the new technology. A perfect example of this is the development of an XML version of the MARC record. There are indeed similarities between MARC and XML, primarily that both can be used to mark up or encode machine-readable documents. Both can also encode structured data, although the MARC use of fixed fields is less flexible than XML, which allows variable-length data throughout. MARCXML was developed as a pure serialization of the MARC format. "Serialization" means that the data encoding of MARC was translated directly to XML without any related transformation of the data itself. Although this produced a record that could be managed with XML-aware software, it did nothing to improve the kind of data that could be conveyed in library bibliographic records. It also did nothing to address some of the limitations of the MARC record. The MARCXML standard is kept one-to-one with the original MARC record, with the single exception that field and record sizes are not enforced. (MARC fields are limited to a four-character length, thus to 9,999 bytes; the record itself cannot exceed 99,999 bytes.) But the limitation on the number of subfields to a field remains, even though there are fields that have no open subfields available for expansion. Other inconveniences also remain, such as the non-repeatability of the MARC fixed field information, which then forces some repeatable elements like languages and dates to be coded in more than one field to accommodate repeatability. MARCXML was never allowed to develop as its own technology, and therefore did not present a change. Library data in XML, rather than in MARCXML, could have represented a real change in capabilities. It might also have provided a better transition to new technology than we now have, because we could have resolved some of the more awkward elements of MARC over a decade or more, with a gradual update to the library systems that use this data. Today we either have to carry those practices on to our future data, or we need to make a great leap forward and break with our past.

We missed the XML boat, but now some are hoping to get on board the latest ship sailing by: the Semantic Web and its base technology, the Resource Description Framework (RDF). It should be noted that there is one other data technology development that could have been considered between XML and

RDF, that of object-oriented design (OO). By the early 1990s, when the FRBR Study Group was being formed, relational technology was no longer new and object-oriented technology was taking its place in many implementations. Programming languages like Java and Python are object-oriented, and data and databases can also be "OO." Library data is leap-frogging over this technology, or it will if it adopts RDF for its data, as it appears it might.

Unlike most of the data models that preceded it, from entity-relation to object-oriented, RDF does not arise from the world of business that prompted our previous technology upgrades. The Semantic Web, as the name implies, comes out of web technology. This is a significant difference from, for example, database technologies, because the web is an open platform and is the place where we put publicly accessible data, whereas databases are private and closed, housed within enterprises and often highly controlled in terms of access. This means that many of the design assumptions that drive the Semantic Web standards are quite different from those encountered in business data processing.

First, let's look at where the Semantic Web comes from and what is meant by "semantic." The Semantic Web comes out of a combination of web technology, with linking and identifying as primary requisites, and the artificial intelligence (AI) community, with smart "bots" as its goal. Where most of us read the term "semantic" as meaning "meaning," in the AI world "semantic" refers to a computable axiom, such as:

If A = B, and B = C, then A = C

Obviously, machine intelligence and human intelligence are significantly different. AI attempts to model human thinking by defining the world as information about things and rules that can be used to "understand" those things. As we know from the overly confident promises that have come out of the AI community since the dawn of computing, the world and how we humans understand it is more complex than it seems. Human intelligence is still a marvel that is unchallenged by machines, in spite of gains in such algorithm-rich areas like the game of chess.

Artificial intelligence on the World Wide Web is a more tractable problem than creating a robot that can navigate stairs, recognize human faces, and pass a Turing test, because the web is already a data abstraction with some distance from the sense-experienced world, and therefore more amenable to computation. The Semantic Web was introduced in an article in *Scientific American* in 2001 by Tim Berners-Lee, founder of the World Wide Web and director of the World Wide Web Consortium, and his associates James Hendler and Ora Lassila. The article told the story of a helpful bot that could find an available doctor, check

your calendar, and make an appointment that fit into your schedule. Creating such technology over the web would require much less effort than creating this technology as a stand-alone system; the web already had solved the problem of a large distributed system capable of handling heterogeneous data and billions of users. The trick was to include in the web the kind of coding that would allow data to be used alongside the current web of documents and media files.

The technology to achieve this is all based on the Resource Description Format (RDF), which itself is a deceptively simple model of things and relationships that can be used to express very complex data. There are some particular aspects of RDF that are both essential and notably different from the technology that most of us have worked with during our careers. There isn't space here to fully elucidate the technology that is RDF, but some points are key to the analysis in the second part of this book. Let's begin with identifiers.

Everything being described in RDF must have a standard identifier that begins with "http://" followed by a domain name (e.g., "ala.org/") and a precise path (conference2015). That might seem confusing, because that is the same prefix that is used for a uniform resource locator (URL), which is the address of something on the web. RDF is using the same standard for its identifiers for a couple of reasons: first, the mechanism to create and manage domain names on the web already exists, which means that it will be easy to create these identifiers; second, the combination of identification with location means that information about the thing identified can be stored on the web at that location without any change in technology.

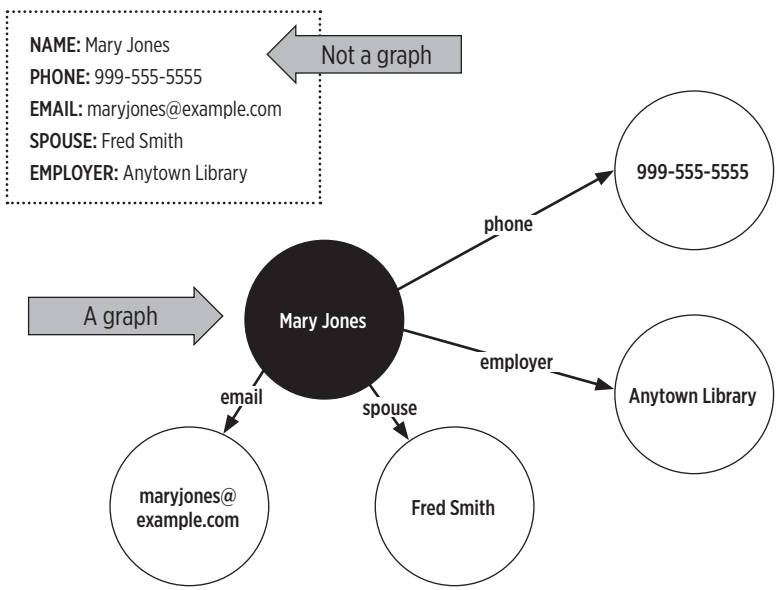
RDF identifiers are intended for machines, not humans. No one wants to read, much less type, "http://id.loc.gov/authorities/subjects/sh85038796" for the Library of Congress subject heading "Dogs." All identifiers can have human-readable labels, and the assumption is that in every situation where a human is interacting with the data, the human-readable label will be the one displayed. This includes input, which in many data creation scenarios in business applications already makes use of textual pull-down lists for easy and accurate input. Thus a cataloger will choose a subject heading, such as "Dogs in literature," from a list and the data stored will be "http://id.loc.gov/authorities/subjects/sh85038823."

Identifiers are in a sense merely a substitution of the normalized text we use today, often in the form of a formatted heading, with a particular string in the URL format. Other changes required in the shift to RDF are more radical. One of the ones that is most difficult to understand is that RDF data about resources is not stored as separate records; instead, information about a thing is in the form of a graph of statements. Graphs have no boundaries; they can grow and they

can interconnect with other graphs where their data intersects (figure 3.1). To give a simple example, the identified author in a library catalog description can interlink with the author information page on Amazon or with the encyclopedic entry about the author on Wikipedia. This assumes that these systems have knowledge of each other's identifiers, but that is increasingly the case: library authority identifiers are already found in Wikipedia entries, so this connection can be made. Data in RDF resembles synapses, with multiple connections that allow new information paths to be created as more information is added (figure 3.1).

FIGURE 3.1

A graph



The next key piece of information about RDF is actually about the nature of the World Wide Web itself. The web is an open space where millions of people and corporations and governments can put information that they wish to make public. Most contributors to the web also have other information stored in private data repositories. Although these private repositories may be in some way connected to the Internet, they are protected by user accounts and passwords, and some are protected through layers of digitally locked doors. The Semantic Web has an emphasis on the public information space, although its technology can also be used for privately held data.

There are three main principles that govern the Semantic Web that are important for understanding the rules that are applied to Semantic Web data:

- the Open World Assumption
- the Non-Unique Name Assumption
- "anyone can say anything about anything"

The Open World Assumption describes the nature of the web, which is that the web is never complete, never done, and it may not be possible to have access to all of it at any one given time. What this means is that web applications must not rely on completeness. If your bibliographic description on the open web has no title, it doesn't mean that there will never be a title, or that there hasn't ever been one. You can assume that a title exists, just not in your current view. Contrast this to a database application that has strict control over input and output, and where rules governing the data are enforced: that title must be there. In a database, a bibliographic description with no author means that the resource has no author attribution. In the web environment, that negative cannot be assumed from the absence of the element.

The Non-Unique Name Assumption (NUNA) states that any identified thing can be identified with more than one identifier. This is like real web life, where I am identified by more than one e-mail address (one at kcoyle.net and another at gmail.com), an IRC handle, and a Twitter name, in addition to my social security number, passport number, driver's license number, and so on, in "real life." On the web you cannot assume that each identifier represents a unique entity. To avoid chaos, there are ways to code identifiers as identifying to the "same" or "different" resources, but the Non-Unique Name Assumption rules any identifier pairs without explicit relationships, such that you cannot draw conclusions from identifiers alone.

The statement that "anyone can say anything about anything" is as true for today's World Wide Web as it is for the Semantic Web: there is no technical restriction on who can put information on the web. There is also no restriction on who can link to resources on the web. You may exercise content control over a web site that you create, but you cannot stop anyone else from linking to it. The same is true on the Semantic Web, where anyone can create links to your data. There is, however, a difference in the effect of linking on the Semantic Web as compared to the web of web pages, because RDF links are more meaningful than links between web pages. Links between web pages have a single meaning, which is simply "this links to that." Semantic Web links carry a meaning to the link,

such as "this is a sub-class of that," or "this is the same as/different from that." These are conditions that you should keep in mind when designing your data. To the extent that you can predict how your data might interact with other data in that vast data space, you need to design your data to "play well with others."

The basic technology of the Semantic Web is RDF. Other technologies build on that. One of these is the Web Ontology Language, OWL, which is the language developed for the creation of Semantic Web vocabularies. First, yes, OWL should be WOL, but it is OWL. Second, the RDF documentation uses the terms *vocabulary* and *ontology* interchangeably. The term *ontology* comes out of the artificial intelligence community and it implies a level of rigor in the definition of terms and their relationships. OWL is to the Semantic Web what a metadata schema has been for us in the past: OWL is how you define the terms of your domain and how you will use those terms to create your data.

OWL is a difficult standard to understand if you are not familiar with certain aspects of artificial intelligence decision-making. Many of the features that are defined in OWL sound familiar but in fact mean something different from what most of us are accustomed to. OWL is designed for a particular Semantic Web function called "inferencing." Inferencing allows you to draw conclusions from data that is present. Thus if:

Every man is a mammal
Fred is a man
Therefore, Fred is a mammal

OWL is quite a bit more sophisticated than this example implies, and includes concepts such as "inverse functional object property" and "negative data property assertion," among many others. The purpose of OWL is to define a vocabulary that can be used in complex artificial intelligence work. It also includes the ability to define some common features of metadata languages, such as cardinality (mandatory, repeatable) and equivalence (same as or different from). Unfortunately, what these features mean in OWL can be quite different from what they mean in metadata standards with which we are familiar.

The meaning of the OWL terms is governed by the RDF concept of classes, in which things being described acquire their membership in a class from the terms that define them. In our simple example above, Fred acquires "mammal-ness"

because he is described by the term "man," which itself has been defined as being of class "mammal." In artificial intelligence this mimics the human brain's ability to draw conclusions from information in the environment, generalizing from knowledge gained in one experience to apply in other situations. The Semantic Web builds up knowledge from atoms of learning, which is the opposite of the top-down approach that is common in classifications of knowledge.

There have been controversies about OWL since its inception, because it is so very complex and also so easily misunderstood. Depending on your application, you can ignore much of that complexity, but for any OWL assertion that you do use you must make sure that you understand the consequences of its use. In particular, many of the OWL declarations about terms and classes seem identical to functions in familiar programming languages. A simple example mentioned above is cardinality. Cardinality in programming languages declares the minimum and maximum allowed occurrences of a data element. If the minimum cardinality of the element is "1," that element is required—it must occur at least one time. If it is "0," then the element is optional. If the maximum cardinality is "1," the element is not repeatable, but any other number defines the number of times it can repeat in your data. In most programming situations, data that violates these rules is considered to be in error.

OWL has minimum and maximum cardinality, but their meaning has a different interpretation due to the application of the Open World Assumption and the Non-Unique Name Assumption. You can define your data as having, for example, a single creator for each given resource; the maximum cardinality of your creator element is therefore "1." If you create or encounter data that has more than one creator for a single resource, this is neither an error nor even an inconsistency in the data. Instead, applying the rules of the Semantic Web, applications that interpret OWL data will conclude that all of the creator identifiers identify a single entity because your rule says that there is only one such entity, and that entity can have any number of identifiers. At times this OWL rule may come in handy because you want to find equivalent identities, but that presumes that the data has all been coded correctly, something that most of us have learned is rarely the case. This is the big "gotcha" of OWL. OWL-based software can examine data that exists and can return a response that the data either does or does not conform to the OWL rules that have been defined for those data elements. But OWL cannot control the creation of data that meets its rules; it examines but that it does not enforce, in large part because "anyone can say anything about anything" and because OWL is intended to function in an open world that is always in flux.

This aspect of OWL generally confuses people because the OWL rules so closely resemble the rules that other programming languages use for a very different purpose: data quality control. In fact, because people often want to use OWL rules in the same way that they use programming rules in closed and controlled environments, there is now software that applies the OWL rules in closed environments, treating identifiers as uniquely identifying a single entity. This reverses two of the main truths of the Semantic Web, which are the Open World Assumption and the Non-Unique Name Assumption. It also operates on data stores where "anyone can say anything about anything" is definitely not allowed. In other words, a mirror copy of the OWL language is being used in the same way that we have always used programming languages, but not in the way intended for the Semantic Web.

Within your own closed environment, such as a local database, you clearly can do whatever you want with your data and you can impose any kinds of rules and controls that serve you and your organization. But if you open that same data to the web, the meaning of those rules will be interpreted using the Semantic Web standard meanings, which means that the Open World Assumption and the Non-Unique Name Assumption will be applied. The actual meaning of your data will be radically different in those two different environments, and operations like searches could yield very different results. The upshot of this is that the same OWL-defined vocabulary should not be used in both the closed and the open worlds.

This conflict between the controlled data stored in one's personal or corporate database and the open environment of the web is one of the hardest for data designers coming from other technology environments to overcome. There obviously is a real need to perform quality control on data, but the basis of the Semantic Web is one of discovery, not control. This is a conflict that, as of this writing, is unresolved, both in code and in terms of best practices. One possible solution, proposed by the Dublin Core Metadata Initiative (DCMI), the same people who develop the Dublin Core metadata terms, is to separate the controlling aspect of the vocabulary from its basic semantics. This isn't different from many existing metadata implementations: terms to be used are defined for their meaning, and a separate structure and rules are developed that turn those terms into a metadata record.

Dublin Core (DC) is a good example of this. Dublin Core terms are defined apart from their use in metadata. Dublin Core's element "title" is defined simply as "the name of the resource." Whether it is mandatory or optional, and whether or not it is repeatable, is not part of the definition of the term itself. Those rules would be defined in a metadata schema or in what the DCMI calls an

"application profile," which is a definition of the metadata structure and rules for a particular application. The term can be used in different ways in different metadata implementations, and the DC terms are indeed used in a wide variety of situations. However, in all uses the term retains the same meaning. This separation of meaning from rules results in maximum flexibility that allows the same terms to be used in many different applications, as Dublin Core terms are today. That flexibility is the positive outcome of this method. The negative outcome is that the separation of meaning from rules results in maximum flexibility, so that data sharing requires some adjustment between communities. The application profile, if provided in a machine-readable form, can be the basis for data sharing because communities can easily understand the structure of data created by others. Through all of that, however, a Dublin Core title remains "the name of the resource" even if some communities allow only one, some more than one, and for some the element may be optional.

We can contrast this to the primary metadata standard used in libraries today, MARC 21. This standard defines the meaning of terms and also the rules for data quality in a single standard. This is not uncommon as a data creation and management approach, however, it is undeniably a definition of a closed data world. Anyone who would use the base MARC record structure and data elements with a different set of rules governing term meanings and cardinality would simply not be creating MARC 21 data, and there would be no expectation that one could successfully combine data created under such different sets of rules.

The final aspect of the Semantic Web that I'll cover here is classes. We're all familiar with the concept of a class from scientific taxonomy and classification systems. In those systems we assign things to classes to give them the meaning of the class, putting ourselves and cats in the class "mammals," and books on mammals in one of the sub-classes of biology. Classes have a different meaning and work differently in the Semantic Web; they are not categories or boxes to put things into, but are meaningful information about things that can be used in various contexts. Classes are not exclusive in their nature, and anyone or anything can have the qualities of more than one class. This is much like the real world, where a person can be an employee in one context, a parent in another, and a volunteer firefighter in yet another. Rather than assigning a thing to a class, the class is deduced based on how something is described. Our rules may say that persons with paychecks are employees, those with children are parents, and those who are members of Volunteer Brigade 7 are volunteer firefighters, and anyone can be all three. By attributing characteristics to the thing we are describing, we build up our world by describing it. This, too, fits into the methods of artificial intelligence where their creations must be able to make deductions about newly

encountered things in the world based on information, as we do in real life. We recognize chairs as chairs even if we haven't seen a particular chair before. We understand that a person is a police officer because anyone wearing that uniform is a police officer, even if we haven't seen that person before. We are moved to open the door for a person carrying packages because we know that it's hard to open a door when your hands are full, in spite of not having been in this exact situation (same person, same door, same packages) before. All of this computation happens quickly and naturally in the human brain, and some of it can be imitated through code if the right information is given about things we describe on the web.

The preceding describes some of the fundamentals of the Semantic Web. The Semantic Web is implemented as *linked data,* a set of common practices for data on the web. One of these practices, the use of http-based identifiers, has been discussed above. Other practices have to do with making sure that your data can be used in the open environment of the World Wide Web. There are standard ways to define your metadata so that others can understand it and potentially use it. Linked data is a mix-and-match technology, and people are encouraged to make use of metadata definitions that exist rather than inventing their own. Any description can be made up of metadata from a number of different sources, and can use descriptive elements found anywhere on the open web.

From this description you can undoubtedly conclude that a future library data standard using linked data would look considerably different from the data we have today. The purpose of linked data is both discovery, through hyperlinks, and new knowledge creation, by linking between previously separate communities and their data stores. Those looking at linked data for libraries are focused on the library catalog and its discovery function. Our current catalog data is very different in its goals and content from data that would play well in a linked data environment. The challenge for us is to make this transition intelligently, and in a way that serves library users. The remainder of this book looks at current efforts with that challenge in mind.

PART II

FRBR AND OTHER SOLUTIONS

FOUR

FRBR IN CONTEXT

Nearly twenty years after the first draft of the Functional Requirements for Bibliographic Records (FRBR), and fifteen years after its final draft was released, the use of concepts and structures defined in FRBR is assumed to be the way forward. Yet it isn't at all clear to me that we have an understanding of what FRBR means for library practice and for library users.

The FRBR model of bibliographic data is the most radical change to library catalog thinking since Panizzi developed his ninety-one rules for the British Museum Catalog. The model presented by FRBR is complex, but it is made even more complex by the competing concepts in the Final Report from the FRBR Study Group that developed FRBR (IFLA 2009). FRBR is alternately seen as an analysis of user needs, a description of the cataloging workflow, and as a data model for a future bibliographic record format. It is rarely viewed, however, as what it was originally intended to be: as the development of basic requirements for an international standard national bibliographic record.

Few of us have done a close reading of the Functional Requirements for Bibliographic Records document, although I would guess that many have glanced at the diagrams, either within the context of the document or as illustrations used in presentations that we've attended. This means that many of us have had our concept of FRBR formed from secondary sources that emphasized only a portion of the content of the document. The FRBR Final Report is 142 pages in length, including appendixes, which makes it a formidable read. The document includes three very high-level entity-relation diagrams—high-level in the sense that they contain very little detail. Although a picture may be worth a thousand words, these three diagrams are far from expressing the full meaning of the work of the FRBR Study Group. There is some ambiguity between the textual description of the conceptual model and the entity-relation diagrams that have come to represent FRBR for most librarians.

Most discussions of FRBR begin with a list of the entities in the three groups, and then illustrate these entities with one or more of the diagrams from the document. I hope to do something quite different here, which is to focus on the text itself, and how the text describes goals and conclusions of the study. I also will provide some historical context for the work of the Study Group.

With the implementation of FRBR concepts in the cataloging standard Resource Description and Access (RDA), FRBR has been accepted by some members of the library community. However, there have been few studies testing the library user's view of FRBR and there are not a great number of implementations of FRBR as an actual catalog.

A report produced in 2006 for the Library of Congress by Karen Calhoun recommended investigating FRBR, which showed that the author did not consider FRBR a "given," but only one possible direction for bibliographic data: "4.2.6 Support experimentation with FRBR and urge vendors and library service organizations to implement clustering based on FRBR concepts" (Calhoun 2006, 18).

The 2008 Library of Congress report on the Future of Bibliographic Control, titled *On the Record* recommended that all work on RDA be halted while studies can be done on the viability of FRBR. Although these two major LC reports called for systematic investigation of the ideas presented in FRBR, that did not happen In the meanwhile, FRBR concepts were incorporated into RDA, which had an implementation date of March 31, 2013, for participating libraries.

The analysis in this book takes a broad view of the cataloging culture that preceded FRBR in an attempt to understand the motivation of the members of the FRBR Study Group. What problems were they trying to solve, and what were the tools at their disposal?

We will see that the primary direction taken by FRBR Study Group, using an entity-relation analysis model, greatly influenced the outcome of the study. Proponents of that method in the FRBR Study Group were also key members of the cataloging standards community that developed the successor to AACR that began shortly after the publication of the FRBR Final Report. After a false start on AACR3, the cataloging rules were reborn as rules for an implementation of the FRBR entity-relation model, RDA.

It is clear that these two standards, FRBR and RDA, were heavily influenced by the thinking of a small group of people, perhaps no more than a score. Even if the meetings of the FRBR Study Group were open to the public, the standard was developed by a group with a closed membership and who did not use available social media to extend the conversation and deliberation beyond itself. Comments were solicited from IFLA institutional members, but, of course commenting on a draft of a document is far from participating in its creation. That said, this is a common standards development method in the library community, which contrasts with the groups that develop the Internet and the World Wide Web. The Internet Engineering Task Force allows anyone to make proposals relating to the technology. Those proposals are called (and remain throughout their use) "Requests for Comments." Changes are not only discussed; they are implemented in code as the proposal progresses through discussion. Anyone can participate in the development of standards. The World Wide Web Consortium, another standards development body, does have members—over 400, in fact. Members are companies and institutions. Some committees are limited to member representatives, but most communication takes place on open mailing lists to which anyone can post, and document drafts often are developed on publicly accessible wiki pages. Members can submit documents that discuss or propose web-related technology.

Another significant difference between these library standards and standards in other communities is that library standards not only do not provide proof of concept through "running code," they actually eschew technology altogether. At least, they claim to. Both FRBR and RDA are stated to be "technology neutral." This is obviously not true, because the analysis in FRBR made use of a very specific technology, the entity-relation model. Perhaps it would have been more accurate to say that FRBR was "application neutral." However, it is probable that the members of the group did not understand how much the technology of relational model determined the group's outcome. As we'll see, there is at least some evidence that the entity-relation model was not well understood, and that this has resulted in some contradictions between statements in the text of the report and presentation of the model as entity-relation diagrams.

What is notable about FRBR, and in some respects RDA also, is that it makes numerous assumptions that were never tested. Because FRBR was couched in terms of a known technology, it was assumed to be technically valid and perhaps even implementable, in spite of the declarations of technology neutrality. Yet no implementations of FRBR, even on a small set of test data, were developed as part of the FRBR Study Group's process. RDA is therefore a cataloging standard based on an unproven conceptual model. The technology that would support them is, at the time of this writing, still unavailable.

In spite of lack of proof of FRBR as a bibliographic model, the concept of FRBR has reached beyond the library community. These implementations often differ considerably from the presumed library implementation. Unfortunately, these variations generally do not provide an explicit statement of their interpretations of FRBR or why they chose a different reinterpretion of FRBR as defined in the FRBR Final Report.

To understand how the library community got to this point, it is necessary to revisit the context in which the FRBR standard was developed.

FIVE

FRBR
STANDARD FOR INTERNATIONAL SHARING

As I write this in 2015, the term *FRBR* has meaning on its own. The full form, Functional Requirements for Bibliographic Records, is still known but has faded into the background. The concept that that library world needed functional requirements for its bibliographic records was foremost in 1992, when the Study Group on Functional Requirements for Bibliographic Records was formed as an IFLA group. We must go back to that time to understand the original intention of the task, and what problems it was asked to solve.

The impetus to define functional requirements for bibliographic records (FRBR) arose out of the IFLA-sponsored Stockholm Seminar on Cataloguing in 1990. The IFLA cataloging section had been addressing international cataloging standards for decades, most prominently through the creation of the International Standards for Bibliographic Description (ISBD) and the International Cataloguing Principles (ICP1961). I don't find evidence that the seminar produced any of its own documents; 1990 predated the IFLA web site, which is where outcomes are reported today. I therefore rely on reports written at later dates. Significantly, the key reports

were written by American members of the FRBR Study Group, and therefore may not reflect exactly the point of view of other participants at the Stockholm Seminar.

Barbara Tillett's 1994 report on the meeting and its outcomes describes the factors that led to the assignment of the FRBR Study Group; she refers to "the mounting costs of cataloging," the proliferation of new media, "exploding bibliographic universe," the need to economize in cataloging, and, "the continuing pressures to adapt cataloguing practices and codes to the machine environment." Regarding the concerns about the costs of cataloging, she states: "Some speakers proposed that cataloguing could be considerably simplified. One speaker stated that the number of descriptive data elements needed in a bibliographic record could be reduced without seriously affecting access" (Tillett 1994). In that same document, Tillett states that the members of the Seminar could not reach consensus on the "functions of bibliographic records," and failed to reach "common agreement on what the bibliographic record is to achieve in answering user needs." This is a strong statement about the perceived state of cataloging in 1990, 150 years after Panizzi drew up his ninety-one rules, 110 years after Cutter's statement of the objectives of the catalog, fifty years after the first International Standard for Bibliographic Description, and over two decades after the creation of the detailed cataloging rules in the Anglo-American community.Yet this harks back to Seymour Lubetzky's 1946 criticism of the cataloging rules: they appeared to be arbitrary because they did not include a functional justification for the purpose of each rule. For catalogers, it might have been disconcerting to discover that the rules that they had been applying for years did not have any specific user goals behind them.

Olivia Madison, who chaired the FRBR Study Group for part of its time, reported on the group's results at a meeting of the International Congress on National Bibliographic Services in Copenhagen in 1998. Her summary of the activity stated: "The central goals of this study were to assist in decreasing the costs of cataloguing by encouraging the sharing of bibliographic data records and to recommend the most useful and important data elements in those records for their users" (Madison 1998).

The goals, then, that prompted the formation of the FRBR Study Group were:

- Determine a minimum set of data elements needed to satisfy user needs.
- Reduce the costs of cataloging.
- Encourage sharing of bibliographic records (internationally).

It is interesting to compare these to the goals of ISBD. The ISBD document opens with this statement of its goals:

> The primary purpose of the ISBD is to provide the stipulations for compatible descriptive cataloguing worldwide in order to aid the international exchange of bibliographic records between national bibliographic agencies and throughout the international library and information community. (ISBD 2011, 1)

One would think that "basic bibliographic data" and "consistency when sharing" would fit in with the goals that came out of the Stockholm Seminar. Given the proximity of these goals to those of the FRBR study, it is not clear why ISBD, or an ISBD variant, was not considered as an answer to the needs as stated, particularly because of the emphasis on "bibliographic records," which is precisely the area in which ISBD performs (and FRBR does not). Comparisons of ISBD and FRBR show how close they are in terms of data elements, which appears to be by design, not by accident. The Final Report of the FRBR Study Group hints at this when it says:

> The attributes defined for the study were derived from a logical analysis of the data that are typically reflected in bibliographic records. (FRBR Final Report, 31)

In terms of defining a "minimum set" of bibliographic elements, by my count ISBD has a few less than one hundred data elements, while FRBR has about eighty-five, which doesn't reduce the number of elements in any significant way.

Some minor adjustments were made to ISBD to avoid conflicts with FRBR, but any other interrelation is difficult to describe. As we'll see, FRBR does not include cataloging instructions nor display information, although both are important aspects of ISBD. In spite of the fact that the term *bibliographic records* is in the name of the standard, FRBR does not address record structure, while ISBD does, although not as the term *record* would be interpreted by a programmer or database designer. ISBD adherence should result in relatively consistent textual output from diverse cataloging departments, and the resulting data, with its specialized punctuation, should be comprehensible even when the language of cataloging is not understood. Because FRBR does not define a record format, there is no equivalent in FRBR to the ISBD punctuation and display rules.

THE FRBR TERMS OF REFERENCE

Coming out of the Stockholm Seminar, a strong motivation for the development of a new bibliographic model was economics: the need to serve users while reducing the cost of cataloging worldwide. The economic issue is addressed repeatedly

in the introductory section of the Final Report of the FRBR Study Group, with statements such as:

> The purpose of formulating recommendations for a basic level national bibliographic record was to address the need identified at the Stockholm Seminar for a core level standard that would allow national bibliographic agencies to reduce their cataloguing costs through the creation, as necessary, of less-than-full-level records, but at the same time ensure that all records produced by national bibliographic agencies met essential user needs. (FRBR Final Report, 2)

A document called the *Terms of Reference*, authored by Tom Delsey and Henriette Avram, gave the official charge to the FRBR Study Group. It defined the problem and the intended outcomes of the group's work.

As defined in the *Terms,* the FRBR Study Group would consist of group members as well as consultants chosen to draft the report. The first consultants were Elaine Svenonius, Barbara Tillett, and Ben Tucker. Over the course of the work, the group of consultants changed. Tom Delsey joined the group a year later when Tucker left. Later, Elizabeth Dulabahn of the Library of Congress was added.

The consultants had four required tasks, as laid out in the Terms of Reference:

1. Determine the full range of functions of the bibliographic record and then state the primary uses of the record as a whole.
2. Develop a framework that identifies and clearly defines the full range of *entities* (e.g., work, texts, subjects, editions, and authors) that are the subject of interest to users of a bibliographic record and the types of *relationships* (e.g., part/whole, derivative, and chronological) that may exist between those *entities.*
3. For each of the *entities* in the framework, identify and define the *functions* (e.g., to describe, to identify, to differentiate, to relate) that the bibliographic record is expected to perform.
4. Identify the key *attributes* (e.g., title, date, and size) of each *entity* or *relationship* that are required for each specific *function* to be performed. *Attribute* requirements should relate specifically to the media or format of the bibliographic item where applicable.

The final requirement was assigned to the FRBR Study Group itself:

5. For the National Libraries: for bibliographic records created by the national bibliographic agencies, recommend a basic level of functionality that relates

> specifically to the *entities* identified in the framework and the *functions* that are relevant to each.

The Terms are more specific than the goals that came out of the Stockholm Seminar; they clearly define how the Group is to go about its work, by defining entities and relationships, a modeling method that had been developed for relational database design. The *Terms* even include a reference to *Data Analysis: The Key to Data Base Design* by Richard C. Perkinson (1984), as background for performing an entity-relation analysis. According to Olivia Madison, who was chair of the FRBR Study Group from 1991 to 1993 and again from 1995 to 1997:

> As mentioned earlier, Delsey had originally authored, with Avram, the initial CDNL [nb: Conference of Directors of National Libraries] *Terms of Reference,* and was well versed in the research literature and potential applications of E-R modeling. In fact, it was largely due to Delsey's commitment to this modeling technique that it was explicitly included in the CDNL *Terms of Reference*. (Madison 2005)

The Terms of Reference were accepted at the 1992 IFLA Conference in New Delhi, after two additions were made by the Standing Committee of the Section on Cataloguing (Madison 2005):

1. Subjects were added to the list of entities.
2. The FRBR Study Group was charged with proposing minimal level standards for bibliographic records.

It is particularly interesting that the first version of the Terms of Reference ignored both subject access (which surely has user implications) as well as the primary purpose of the study, which was to address the costs of cataloging for national bibliographic agencies by creating a minimal-level record requirement. The version of the Terms of Reference published in 1992 makes only a brief mention of the economic goals, referring to "operating under increasing budgetary constraints and increasing pressures to reduce cataloging costs through minimal-level cataloging." The remainder of the Terms document focuses on a technical analysis of bibliographic data.

In the end, what came to dominate the outcome of the FRBR Study Group's work had little to say about minimal level records or about addressing economic issues. The "user tasks" that have become a near mantra in some cataloging circles ("find, identify, select, obtain") are placed adjacent to bibliographic elements in one section of the FRBR Final Report, but their relation to the user is left unstated.

During the development of FRBR there were periods of review, and comments were received, but these have not been made public. (For a profession that prides itself on preserving knowledge for the future, we are terrible at keeping our own history.) Olivia Madison recounts a 1996 review of FRBR that received forty responses from sixteen different countries. Of these, seven were critical of FRBR.

> The principal issues reflected honest difference of professional perspectives related to the research methodology, the actual need for this particular study, the adequacy of its user-focus, and the work process and adequacy of detail. (Madison 2005)

It would be interesting to know which libraries responded, and what those differences of opinion were. That criticism was made of the research methodology (which may or may not refer to the insertion of an E-R analysis into the study) and the adequacy of its user focus seems important from today's perspective. Madison calls these forty responses from sixteen countries a "worldwide review," although it seems a bit "world-narrow." In the end, FRBR was developed and reviewed by a very small constituency that was not representative, by any measure, of the library community that prompted its development. That said, the use of select expert committees to develop standards is not unusual, when few libraries can give staff leave to devote themselves to such demanding activities. While this is understandable, one must acknowledge the fact that the end result of such a process may not actually serve the larger library community.

SIX

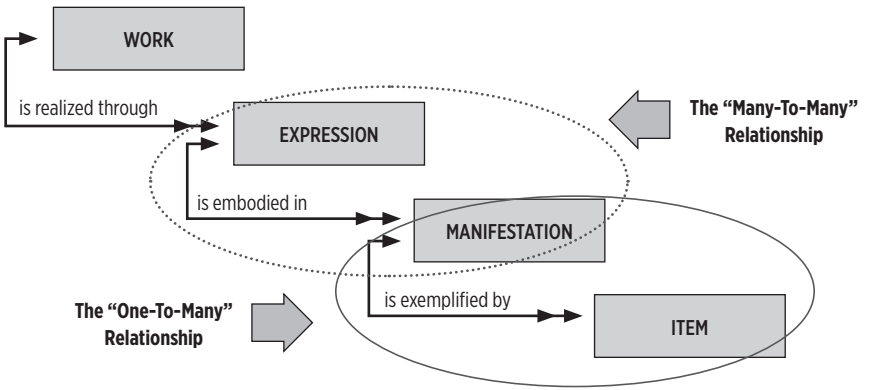
THE ENTITY-RELATION MODEL

Nearly every book, article, or presentation about FRBR has an explication of the primary FRBR entities and their relationships. The entity-relation (E-R) analysis defines the primary structure of the FRBR Final Report, which has chapters for entities, relationships, and attributes. The use of the entity-relation modeling technique was a requirement posed by the Terms of Reference for the study. Because of the great influence that this modeling technique had on the outcome of the study, it is worth examining in some detail.

Entity-relation analysis makes use of particular notation or diagrams to explain what is being expressed. There are many possible notation styles, some using boxes, some ovals, some with many different kinds of arrows and lines, with each notation carrying a specific meaning. The notation used in the FRBR document diagrams dates from the early days of E-R modeling. In this notation there are boxes for entities and arrows for relationships. The arrows can have one head, meaning that only one individual of the entity (e.g., one person) can be related, or they can have two heads, meaning that the relationship can be "many" (figure 6.1).

FIGURE 6.1

Many-to-many and one-to-many relationships in FRBR



Looking at the E-R analysis technique in historical context alongside the work of the Study Group that was developing FRBR helps explain what the diagrams are capable of expressing. Entity-relation modeling and relational databases were developed in the 1970s, and hit their peak in the 1980s. By 1990, the use of relational concepts was overlapping with a new computing paradigm: object-oriented programming and database design. The FRBR group made use of an early version of the E-R modeling concept and notation that was developed in the late 1970s. By 1990, E-R modeling had added design features that allowed the expression of more than just entities and relationships: in these new modeling notations it was possible to indicate inheritance, precise cardinality, processes, and communication paths. These later techniques would have made it possible to indicate whether the FRBR entities were required or optional, something that is not included in the FRBR entity-relation diagrams. Those diagrams show the same relationship between persons and works as between expressions and works, yet we know not every bibliographic description requires that a person be responsible for a work, and the text of the FRBR Final Report states that every expression has a mandatory relationship to a work. These are not distinguished in the diagrams.

By the time the FRBR Study Group provided its first draft in 1994, E-R modeling techniques had been replaced in technical design circles with the Unified Modeling Language (UML), which was developed during the 1980s. UML is a much more expressive language, with fourteen different diagram types, modeling both structures and behaviors. It also is designed primarily for object-oriented analysis, because entity-relation modeling had been superseded by object-oriented design. Had UML been used by the FRBR Study Group the outcome of the

study might have been different, but that also would have required a different skill set on the part of the Study Group members.

For a high-level view, a simple E-R model can still be useful. As its name implies, E-R modeling views one's information domain as entities or things, and defines the relationships between those things. The E-R modeling technique provided a structured approach for the FRBR Study Group, whose task was quite broadly defined. Use of the technique was required by the Terms of Reference document that gave the group its charge. In the Methodology section of the FRBR Final Report, the group explains:

> The methodology used in this study is based on an entity analysis technique that is used in the development of conceptual models for relational database systems. Although the study is not intended to serve directly as a basis for the design of bibliographic databases, the technique was chosen as the basis for the methodology because it provides a structured approach to the analysis of data requirements.

The FRBR Study Group makes clear that the resulting analysis is not a record design, yet there is an acknowledgment that the FRBR Final Report answers some questions that could be applied to bibliographic records:

> The study makes no a priori assumptions about the bibliographic record itself, either in terms of content or structure. It takes a user-focused approach to analyzing data requirements insofar as it endeavours to define in a systematic way what it is that the user expects to find information about in a bibliographic record and how that information is used. The study uses an entity analysis technique that begins by isolating the entities that are the key objects of interest to users of bibliographic records. (FRBR Final Report, 3)

However, the possibility that the analysis could be a precursor to database design was also hinted at in the FRBR Final Report:

> The entity-relationship analysis reflected in the model might also serve as a useful conceptual framework for a re-examination of the structures used to store, display, and communicate bibliographic data.

Barbara Tillett discusses this in her 1994 report on the work of the FRBR Study Group: "We hope this exercise will provide the basics for development of future structured bibliographic databases and future systems that facilitate creation, maintenance, and use of such databases" (Tillett 1994).

At the time, the E-R approach was new to at least some members of the FRBR Study Group and was not part of most catalogers' backgrounds. The FRBR document itself refers to readings in this area that the Study Group members found useful in understanding the entity-relation technology:

> The entity-relationship analysis technique and the conventions for graphic presentation that are used in this study are based in large part on the methodology developed by James Martin and outlined in his book *Strategic Data-Planning Methodologies* (Prentice-Hall, 1982). Graeme Simsion's *Data Modeling Essentials* (Van Nostrand Reinhold, 1994), Richard Perkinson's *Data Analysis: the Key to Data Base Design* (QED Information Sciences, 1984), and Ramez Elmasri and Shamkant Navanthe's *Fundamentals of Database Systems* (Benjamin/Cummings, 1989) were also used in shaping the methodology for the study. All four books are recommended to those who are interested in additional background and more detail on entity-relationship analysis. (FRBR Final Report, 10)

Note the emphasis on database design. Also note the dates on the books cited—the newest is now twenty years old. Relational database design, although still used in business applications, is no longer cutting edge technology. Although modeling of entities and relationships is still common, it has changed significantly from the models in use in these books.

The FRBR entity-relation diagrams show a macro-level model that includes only what are considered to be the primary relationships between entities. Other relationships between bibliographic entities are defined in the text, such as work/work relationships. These do not appear in the diagrams in the FRBR Final Report, thus presenting an incomplete picture of the actual bibliographic model described in the text.

ENTITIES, RELATIONS, AND DATABASE DESIGN

One of the reasons given for the development of an E-R model for bibliographic data was the desire to create a bibliographic data model that was more in tune with current technology. Because MARC was developed as a carrier for printed bibliographic data, and preceded the automation of library catalogs, it wasn't designed with database technology in mind. That doesn't mean that database technology has not been employed in library systems; in fact, they would not function as they do without the storage of data in such systems. Online systems must make use of the efficiencies built into database management systems.

Restructuring MARC data for use in relational databases, as discussed in the chapter on technology, is not an easy task. There are indeed some significant differences between bibliographic data and business data, and there is no question that the main customer base for database technology is the business world, not libraries. Therefore, the database technology that is on the market is optimized for the needs of the majority (and richest) of customers. For example, bibliographic data is primarily textual. Unlike much business data, bibliographic data has few numerical amounts that need computation, and we know that computers are more suited to work with numbers than with text. Also, there is not an even or predictable amount of repetition in bibliographic data; there are some authors or subjects that have high redundancy in a file, but there are even more that exist in a single exemplar. Relational databases are at their most efficient when the same data repeats frequently in the database, but provide less of an advantage for data with a high level of uniqueness.

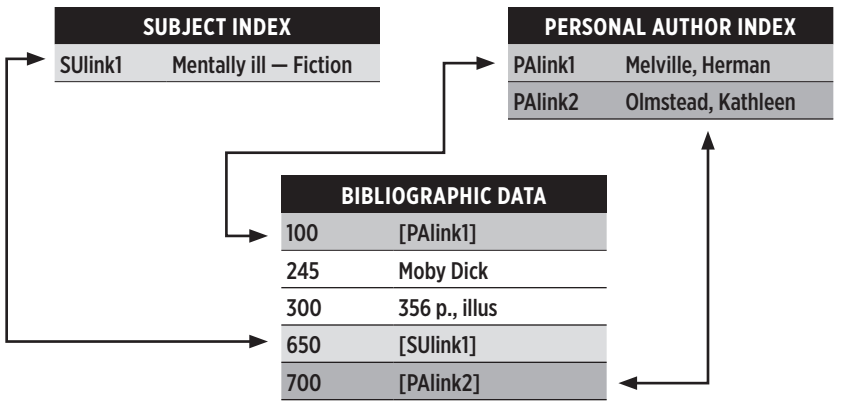
In spite of the fact that bibliographic data isn't what database management system developers had in mind when developing the technology, from the very first every library system has made use of some of the features of a database management system in order to function. It is therefore a misunderstanding to assume that because library data is not easily normalized into ideal relational database forms, library systems do not make use of relational database technology. They do, although the result does not look like the idealized design in database design textbooks.

As part of the development of the FRBR-informed cataloging rules, Resource Description and Access (RDA), Tom Delsey developed the RDA Implementation Database Scenarios, which depict "database structures conventionally used in library applications" (Delsey 2009). Scenario 3 shows library systems working with MARC-based data using a "flat file" approach, which would be similar to data stored in a spreadsheet. In fact, no flat file–based system could produce the kind of retrieval that library systems provide, and most systems today are at least as sophisticated as that document's Scenario 1, which appears to be the preferred model for the management of RDA data. The virtues of the MARC record, with its variability of field and subfield combinations and the unlimited repeatability of most fields, make it unsuited to a flat-file treatment. It would not be possible to provide search or browse on field types, like titles or subjects, without making use of entities and relations. In fact, the database I worked on in the early 1980s definitely used a relational design. A mock-up of its very high-level design is shown in figure 6.2. There were also many other indexes for corporate authors, titles, dates and languages, which I don't show here for reasons of space. An actual

database design is a mass of boxes and arrows that often cannot be reproduced on a single sheet of paper.

FIGURE 6.2

Inside A library system database, circa 1984



There are a couple of points that should be taken away from this. One is that although a data structure that has clear entities and relations defined may be somewhat easy to extend into a database design, many times a database design is derived from data that was initially developed for some other purpose. Although the bibliographic data that is stored in the MARC record still adheres to a structure that originally supported the card catalog, an E-R analysis can be done that results in a database that supports search and display of the data. This database design is primarily based on practical considerations: enabling retrieval of headings and combinations of headings with fast response time, even within large databases. Although the database model for bibliographic data differs considerably from, for example, that of banking, library systems run on the same underlying technology, making use of the features that a database management system provides.

This means that the FRBR E-R model is not the first practical use of E-R modeling for bibliographic data. The opportunity that FRBR afforded was a rethinking of bibliographic data model with entities and relations in mind, which did not adhere to the model of description and headings that has been the form of bibliographic data for centuries. The goals stated in the document— facilitating sharing and decreasing the costs of cataloging on an international scale—may have motivated the FRBR Study Group to develop the entities and relationships in FRBR, although the connection between the goals and the E-R model are not presented clearly in the FRBR Final Report.

SEVEN

WHAT IS MODELED IN FRBR?

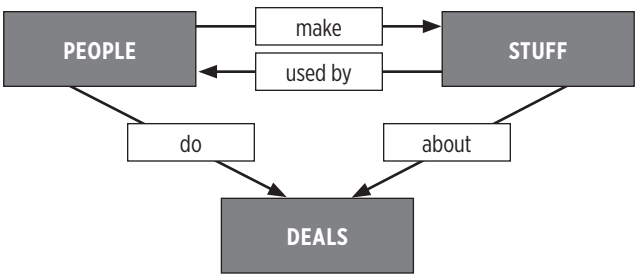
An E-R analysis serves to resolve category boundaries and assign attributes to categories of things or functions. That said, for any given data, there can be any number of E-R models developed, depending on the functionality desired, the requirements of your data management system, and the workflow you need to support. The same is true for bibliographic data. The top-level model developed by the publishing industry has three primary "things": people, stuff, and deals (figure 7.1).

This represents a bibliographic model that primarily supports commercial functions around intellectual resources. The library model developed as FRBR could be described as "people, stuff, and subject access." Each model reflects the needs and views of its community.

The simpler your goals, the simpler your data model can be. However, the FRBR Study Group had a rather complex set of goals. One goal had to do with simplifying the bibliographic record for international sharing, with the purpose of cost savings. Another goal required the Study Group to make a connection

FIGURE 7.1

<indecs> commercial view of intellectual products



between user needs and bibliographic data elements. This was related to the first goal, because the data elements most needed by users would also be the ones that truly should be in the data model. There also seem to have been goals that group members brought to the effort, such as codifying bibliographic relationships between described resources. The use of E-R modeling was itself a goal, which was possibly included because some members assumed that a future bibliographic record would be stored in relational databases, or that sharing would be easier if the data were packaged as separate entities. Because the FRBR Final Report doesn't address technical issues or a record format, those goals are not clarified in the report.

FRBR defines three groups of entities. The groups are not named; they are called only Groups 1, 2, and 3. The groups themselves are included in some of the E-R diagrams as boxes around the entities of the group, but are ignored in further modeling. This means that there are no functions or qualities that belong to the groups themselves. Each entity is treated as separate. There are no group/entity relationships that would create a type of class/sub-class structure. There is also no entity or identified class that represents a whole bibliographic description.

This brings up the question of whether the groupings of entities in FRBR are meaningful at all. Gordon Dunsire, who has created the FRBR representations in RDF for the IFLA FRBR Review Group, appears to have been instructed that the groups are not to be used as classes in the RDF sense.

> FRBR Group 3 is not represented in RDF as a class, following clarification from the FRBR Review Group: the Groups are used to simplify the entity-relationship diagrams, and are not intended to be super-classes. Instead, 10 separate properties are represented in RDF, all with domain Work and each with one of

> the Group 1, 2, or 3 entity classes as range, corresponding to Figure 3.3 in the FRBR report. (Dunsire 2012, 736)

Each of the FRBR groups has a different conceptual structure. Group 1, which could be thought of as representing a bibliographic description, consists of four mutually dependent entities that are modeled as a chain from work to expression to manifestation to item. Although the diagram does not specify whether the entities and relationships are mandatory, it is clear to most readers that all four are needed for a complete bibliographic description, and that having, for example, an expression entity with no work or manifestation entity would not be meaningful. It isn't clear, however, whether the model intends to make all four mandatory as part of a description.

Group 2 consists of *person* and *corporate body,* and these have agent- or actor-type relationships with entities in Group 1. The Group 2 entities, unlike the Group 1 entities, have no relationships that link them to each other analogous to the intra-group relationships of Group 1. They also do not share any attributes. They could have been modeled as members of a class because logically they do share some attributes, like the relationships linking them to the Group 1 entities. Both of these entities could be an author, a publisher, a performer, etc., and therefore those attributes could be assigned to a class that includes all Group 2 entities, but they were not.

Group 3 includes four entities that can have a "subject" relationship to the work entity in Group 1: *concept, object, event,* and *place*. These, too, have no links between them and are not members of a mutual class. Although these four entities are a group called "Subject" in the FRBR text, in fact all entities from Groups 1 and 2 can also have a subject relationship with a work entity. This means that logically all FRBR entities could be sub-classed to a subject class. Group 3 appears to round out the entities needed for subject assignment, but isn't itself a complete list of subject types even though it is referred to as the subject group.

FRBR AS A CONCEPTUAL MODEL

> FRBR is not a data model. FRBR is not a metadata scheme. FRBR is not a system design structure. It is a conceptual model of the bibliographic universe. (Tillett 2005)

The name "Functional Requirements for Bibliographic Records" evokes a much more concrete outcome than was actually presented in the FRBR Final Report.

Functional requirements generally speak to actions, workflows, and methods. A bibliographic record, at least as most of us think of it, is a defined set of data elements that identify and describe a resource. The entity-relation model is also intended to provide a workable model that could eventually be instantiated in some type of computer application. However, the result of the FRBR Study Group's work has often been described as something much less concrete than the name might imply, that is, a conceptual model.

The FRBR Final Report's section 2.3 on methodology gives the reasoning behind the use of entity-relation modeling technique:

> The methodology used in this study is based on an entity analysis technique that is used in the development of conceptual models for relational database systems. Although the study is not intended to serve directly as a basis for the design of bibliographic databases, the technique was chosen as the basis for the methodology because it provides a structured approach to the analysis of data requirements that facilitates the processes of definition and delineation that were set out in the terms of reference for the study. (FRBR Final Report, 9)

Entity-relation modeling is a multistep technique that begins with a high-level conceptual analysis of the data universe that is being considered. To quote once again from the FRBR Final Report:

> The first step in the entity analysis technique is to isolate the key objects that are of interest to users of information in a particular domain. These objects of interest or entities are defined at as high a level as possible. That is to say that the analysis first focuses attention not on individual data but on the "things" the data describe. Each of the entities defined for the model, therefore, serves as the focal point for a cluster of data. An entity diagram for a personnel information system, for example, would likely identify "employee" as one entity that would be of interest to the users of such a system. (FRBR Final Report, 9)

This is a very good description of conceptual modeling. So it is either puzzling or disturbing that most readings of FRBR do not recognize this difference between a conceptual model and either a record format or a logical model. In part this is because it is easy to view the diagrams in the document as statements of data structure rather than high-level concepts about bibliographic data. This may also be because most members of our profession are not familiar with the stages of modeling that are used in formal database design.

One of the common assumptions about FRBR is that the entities listed there should be directly translated into records in any bibliographic data design that intends to implement FRBR. For example, there is much criticism of BIBFRAME for presenting a two-entity bibliographic model instead of using the four entities of FRBR. This reflects the mistaken idea that each Group 1 entity must be a record in whatever future bibliographic formats are developed. As entities in a conceptual model oriented around database design, there is absolutely no direct transfer from conceptual entities to records. How best to create a record format that carries the concepts is something that would be determined after a further and more detailed technical analysis. In fact, the development of a record format might not logically be a direct descendent of the E-R model, because the E-R modeling technique has a bias toward the structure of relational database management systems, not records. In addition, should a conceptual design like FRBR be used to inform the next steps toward a database design, there is no guarantee that the final design will retain the high-level structure of the conceptual model. Few assumptions can be made about the potential technical implementation based on a conceptual design; only further analysis, with a specific technology as its target, can reveal that.

One also cannot make assumptions about record design based on a database design model. For many databases there is no single record that represents all of the stored data. Databases are often a combination of data from numerous departments and processes, and they can receive and output many different data combinations as needed. The database does not define the record format although they must share the definition of the atomic elements that both carry.

The FRBR Final Report has a section recommending areas for further study. In that section the FRBR Study Group states that the report "is intended to provide a base for common understanding and further dialogue, but it does not presume to be the last word on the issues it addresses." Areas for further study include expanding the analysis to authority data, which is being taken up by the ongoing working group within the IFLA Cataloguing Section. The report also suggests performing studies to verify the validity of the attributes listed for the FRBR entities. And finally, it addresses the potential of the entity-relation model to inform a new record format:

> The entity-relationship analysis reflected in the model might also serve as a useful conceptual framework for a re-examination of the structures used to store, display, and communicate bibliographic data. Further study could be done on the practical implications of restructuring MARC record formats to reflect more

> directly the hierarchical and reciprocal relationships outlined in the model. An examination of that kind might offer a new approach to the so-called "multiple versions" issue. The model could also be expanded in depth to create a fully developed data model that would serve as the basis for the design of an experimental database to assess the efficiency and effectiveness of a database structure patterned on the model. (FRBR Final Report, 6)

This recommendation has not had follow-up, and most likely will not, at least not in the sense described here. By the time that the FRBR Final Report was issued in 1998, relational models were on the wane. By 2008, even the successor technology, the object-oriented model, was being supplanted by the data design concepts of the Semantic Web and linked data. Had the library world embraced a relational data design by the end of the 1980s, library data and library systems might have been in line with common information technology development. As it is, the time for a relational design has passed.

GROUP 1
WORK, EXPRESSION, MANIFESTATION, ITEM

Group 1 of FRBR comprises the main entities of bibliographic description: work, expression, manifestation, and item (WEMI). There is a set of relationships between these entities, and they are described in a linear pattern. The FRBR Final Report presents the entities from work to item, moving from the most abstract to the most concrete. One could also take the view of the cataloger's workflow, which begins with the item in hand, and moves through manifestation and expression to work. However, FRBR does not describe the process of creation, but rather a fully realized resource. There is no temporal order implied between the entities of Group 1.

It is the relationships between these entities that complicates Group 1 and also that leads to different interpretations of what possibilities exist to make use of the different combinations of entities. It also creates some complication in the relationships with other entities because each Group 1 entity has separate bibliographic relationships with other FRBR entities. This means that there are not only relationships between work, expression, etc., but there are so-called "bibliographic relationships" between works, between expressions, between works and expressions, and so on. The resulting picture is of a very complex web of relationships.

The FRBR Work

As we saw in a previous section, numerous definitions of "bibliographic work" have been developed in the library field over time. FRBR defines the work entity as "a distinct intellectual or artistic creation." In reference to FRBR work, Barbara Tillett, in her 2003 Library of Congress Pamphlet *What Is FRBR?* defines work as "the conceptual content that underlies all of the linguistic versions, the story being told in the book, the ideas in a person's head for the book." The work as idea is however not entirely borne out by the Final Report text. The FRBR Final Report declares changes in form (e.g., from a book to a movie version of the same story) to be different works, which hints at a more precise definition of work than being a story or idea. This illustrates how difficult it will be to form a widely accepted definition of work. In fact, the FRBR Final Report acknowledges that work may be defined different in different communities.

> Because the notion of a *work* is abstract, it is difficult to define precise boundaries for the entity. The concept of what constitutes a *work* and where the line of demarcation lies between one *work* and another may in fact be viewed differently from one culture to another. Consequently the bibliographic conventions established by various cultures or national groups may differ in terms of the criteria they use for determining the boundaries between one *work* and another. (FRBR Final Report)

The FRBR Final Report also states that "We recognize the *work* through individual realizations or expressions of the work, but the work itself exists only in the commonality of content between and among the various expressions of the work." This is considerably different from the definitions given in the previous paragraph. In this definition, the work is a set of all of its expressions. This would make the work a dynamic entity that changes depending on the presence of particular expressions in the bibliographic universe, similar to Patrick Wilson's bibliographic families that grow as new family members are born. Thus the work is presented both as a fixed a priori abstraction (in the mind of the creator) and as a dynamic entity that is the sum of its expressions.

The work is consistently referred to as an abstraction in library literature about FRBR, but the work as defined in FRBR has the characteristics of something quite concrete: it has a creator, a genre, and subject designation, which is actually quite a bit of "realness" for something that is "an idea in a person's head." It also has a work title, which could be considered especially odd for something that has not be expressed or manifested yet. However, the work title (along with the

creator) serves as the bibliographic identifier for the work, so it has a necessary bibliographic purpose. This is an important point: the definition of the entities in FRBR is not the expression of a philosophical or theoretical declaration of a bibliographic ideal; it is fully grounded in library bibliographic practices, and the model is intended to support those practices.

The FRBR work has many potential relationships with other entities. It has a primary relationship with one or more expressions. (Primary relationships are defined in the FRBR Final Report as the ones visible in the entity-relation diagrams.) It also has primary relationships with persons and corporate bodies. It is the only entity that has a "has subject" relationship with other entities.

The FRBR Expression

The FRBR expression is "the intellectual or artistic realization of a work in the form of alpha-numeric, musical, or choreographic notation, sound, image, object, movement, etc., or any combination of such forms" (FRBR Final Report, 19). The expression is also an abstract entity in that there is no physical realization of the expressed work until it is manifested in a physical format. It has a form of expression (e.g., "text" or "music") and its attributes include a date, a language (for textual works), a medium of performance (for music or other performed works), and scale and project (for cartographic works). The expression excludes, however, "physical form . . . that [is] not integral to the intellectual or artistic realization of the *work* as such" (FRBR Final Report, 19). Therefore, the expression is the artistic realization of work in every aspect except the physical form. It does include "the specific words, sentences, paragraphs . . . or the particular sounds, phrasing, etc. resulting from the realization of a musical work." The expression has all the qualities of a creation, but stops short of any physical attributes. So it is a text, but without a "typeface and page layout." It is music without any readable notes or carrier of sound.

Here we immediately face one of the key issues of the FRBR Group 1 entities: what is the nature of the Group 1 entities in relation to what is essentially a single thing that has been described in four parts? The expression is a realization of a work, and as a realization of a work it not only is not separate from it, it actually is the intellectual content of the work in some form. Although in the entity-relation diagrams and in the lists of attributes the work and the expression are treated as separate, conceptually the expression cannot be separated from the work because the work is the intellectual content of the expression.

The expression is generally considered the most conceptually elusive of the Group 1 entities. Although the work is abstract, it has creative content, creators, and subjects. The manifestation has a physical description and most of us have considerable experience with manifestations in our lives. The expression is described as a text that has no particular physical form, something that seems impossible because just calling it a text must mean that it exists in some form or another.

The expression is also potentially the result of an editorial process or some other third-party contribution, because the FRBR Final Report speaks of expressions with glosses, illustrations, and augmentations. For anyone who envisions WEMI as a directional flow from work to item, or vice-versa, expression is a sticking point because it doesn't have a single direction. The expression is the entity of translations of texts and performance of music, but also of annotated editions and selected works.

It is worth noting here that there is no concept of the creator's original expression in FRBR. All expressions have equal weight. This is potentially an area where Patrick Wilson's second kind of power would come into play, and some qualitative assessment of expressions, along with a family genealogy, could be offered to users seeking the best content for their need.

The FRBR Manifestation

The FRBR manifestation is "the physical embodiment of an *expression* of a *work*" (FRBR Final Report, 20). Like the expression, the manifestation has content that has carried over from the preceding entities. A manifestation may be a unique, stand-alone object, as in an author's manuscript, or it can be a set of mass-produced things made publicly available, or it can be anything in between. The physicality of the manifestation is a primary characteristic:

> When the production process involves changes in physical form the resulting product is considered a new *manifestation*. Changes in physical form include changes affecting display characteristics (e.g., a change in typeface, size of font, page layout, etc.), changes in physical medium (e.g., a change from paper to microfilm as the medium of conveyance), and changes in the container (e.g., a change from cassette to cartridge as the container for a tape). (FRBR Final Report, 22)

This brings us squarely into the "multiple versions" territory, one of the compelling cataloging issues of the 1990s: does a change of physicality require a new

catalog record? The answer in the FRBR Final Report is clearly that a change in physical form is a different manifestation. Therefore a hardback book and a trade paperback of the same book would be different manifestations, and each e-book format would also be a different manifestation. This would be in keeping with the distinctions managed by publishers of these materials, and in the assignment of a specific ISBN to each physical format. However, considering that resources in digital formats, such as electronic books, can be converted to different digital formats at the time of access, it could be said that the physical stability of the manifestation is becoming a thing of the past.

Note that manifestation is not separate from the expression of the work as described here; it *embodies* it. In fact, the FRBR Final Report says that the manifestation embodies an expression of a work, and therefore is defined as the whole Group 1 minus the item. That it *embodies* makes it sound like the manifestation is a kind of container for the expression and the work. That it *manifests* instead sounds more like it manufactures or creates. In addition, the manifestation contains certain elements provided by the publisher, like a title page, often an index, and in some cases some prefatory material, which then brings into question if or when the publisher's additions constitute the creation of a new expression. In some instances, any one particular expression is only one of the creations that is manifested in a publication. Such publications are referred to in FRBR as "aggregates." Although addressed in the FRBR Final Report, aggregates pose some difficult problems and had to be addressed by a special sub-group of the FRBR Study Group. At this point it should be obvious to the reader that the four boxes and few arrows of the diagrams in the FRBR Final Report are a gross simplification of the bibliographic domain that FRBR attempts to capture.

The FRBR Item

Item is the most neglected of the Group 1 entities in most discussions, possibly because it is also the most clear. "The entity defined as *item* is a concrete entity" (FRBR Final Report, 24). An item *exemplifies* a single manifestation. However, even here the meaning has some caveats. An item may consist of more than one physical object, such as a multivolume monograph, or a film or sound product issued on multiple disks. Each item exemplifies a single manifestation. In certain cases, such as with copies of special interest, there may be some variation between a single item and other items of the same manifestation, such as books that are signed by the author, or items that are damaged.

GROUP 2
PERSONS, CORPORATE BODIES, FAMILIES

In the FRBR Final Report, Group 2 has two entities, person and corporate body. Subsequent work on the Functional Requirements for Authority Records (FRAD) has added a third entity, *family,* which is now assumed to be included in FRBR. The FRBR document does not add any new features to persons and corporate bodies compared to current authority records, although FRAD does expand the attribute list to include others, such as gender and places of birth, death, and residence for persons, and language, history, and field of activity for corporate bodies.

As with the other groups, there is no whole that would unify Group 2 as a meaningful unit, even though the entities are treated as a single unit in the text. Without a more general class or entity to which these could belong, some characteristics, like roles, will logically need to be defined separately for each entity, with a significant amount of redundancy. For example, both persons and corporates bodies can be publishers, collectors, dedicatees, defendants, and so on. The lack of a super-class that unifies Group 2 means that it will be necessary to decide which roles are appropriate for each Group 2 member. As we will see, those implementing FRBR as a conceptual model usually develop a super-class for the entities in this group, which also enables working with data where the identity of the agent (especially person or corporate body) has not been clearly determined.

GROUP 3
SUBJECTS

The nature of the third FRBR group is something of a miscellany: concept, object, event, and place. It isn't made clear in the text how FRBR Study Group members arrived at these entities. They do not match the elements of the Library of Congress Subject Headings, which are *topic, form, chronology,* and *place*.

While Group 3 is known as the subject group, it is not the full range of subjects in FRBR. All of the FRBR entities can be in a subject relationship to the work.

> The diagram also depicts the "subject" relationships between *work* and the entities in the first and second groups. The diagram indicates that a *work* may have as its subject one or more than one *work, expression, manifestation, item, person,* and/or *corporate body.* (FRBR Final Report, 17)

Robert Maxwell (2008) notes in his book on FRBR that the limitation of subject relationships to the work entity ignores some cases in which expressions or even manifestations can have a need for topical access on their own. He suggests that "it might be preferable in FRBR simply to define Group 3 entities as 'entities that serve as subjects of Group 1 entities'" (Maxwell, 15). Actually, that definition would also have to be applied to Groups 1 and 2, because all entities can be the object of a "has subject" relationship.

We don't learn any more about the Group 3 because the only attribute provided for each of the entities is "term." The document expressly limits the use of the Group 3 entities to subject access, although in the introduction there is an acknowledgment that an entity like "place" might have use within the bibliographic description, because it would also be logical for an object to have a place, and for that place to have a name and a geographical location with longitude and latitude.

We also learn little about the use of these entities, because other than defining these entities and showing an E-R diagram with the subject relationship, subjects are not mentioned in the document. Quite astonishingly, they are not listed in the discussion of the user tasks. The find task lists:

> the attribute or relationship is typically used as a primary search term for locating the entity (e.g., the title of a *manifestation,* the relationship between a *work* and the *person or corporate body* principally responsible for the *work*). (FRBR Final Report, 84)

We should remember that subjects were not initially included in the Terms of Reference, but were added to the list of entities when the document was reviewed at the IFLA meeting in 1992. It is evident that there is a strong separation between descriptive cataloging and subject cataloging in library practice, and this has an effect on the FRBR Study Group's attention to the subject aspect of user needs. This points out a chasm between library cataloging and user service that FRBR does not address.

THE BIBLIOGRAPHIC RELATIONSHIPS

Nearly all attention on FRBR today focuses on the ten (or eleven, if you count family) entities. Little attention is given to the bibliographic relationships that FRBR defines. Although they echo bibliographic relationships in current cataloging, in FRBR each relationship is designated as being between specific FRBR

Group 1 entities. These bibliographic relationships are distinguished from the primary relationships that exist between the FRBR entities, and which are shown in the diagrams in the FRBR Final Report. The bibliographic relationships, when added to the basic FRBR entity-relation diagram, create a very complex web of connections.

Both Barbara Tillett (1988) and Richard Smiraglia (2001) conducted quantitative studies of the relationships that exist in library catalogs. Tillett did her study on a large portion of the Library of Congress MARC file. Smiraglia researched the catalogs of a small number of research institutions, including two specialized collections, one of theology and one of music, and focused on works and work families (the latter included many relationships that FRBR would define as between expressions). Both found a high number of related resources in the catalogs they studied, although exact numbers are difficult to assign because making bibliographic relationships manifest is optional in library cataloging. Relationships are generally provided only in notes, and only if the cataloger deems that the note is needed for clarity.

It isn't clear if the presence of bibliographic relationships in FRBR implies that they are to be coded as actual links between descriptions, or if they are simply a conceptual rendering of the idea of bibliographic relationships. If one decides that the relationships should be treated as links between entities, then the optionality of explicit relationships described in the report becomes problematic because not all related entities would be linked. In section 7 of the FRBR Final Report, where the basic level record is defined, each of the notes that defines a bibliographic relationship has a footnoted explanation to the effect that such notes are made only if the relationship cannot be inferred from other information in the record. However, if the relationship, as implied in the use of an information technology model for the entities and relationship, is intended to be actionable in a database or other information system, there is no means to "infer" relationships; all data to be acted on must be explicitly coded. Thus, the assumption that FRBR represents a machine-actionable model is clearly contradicted in this area of the FRBR Final Report, leaving us to wonder exactly how we should interpret the use of the entity-relation model vis-a-vis the catalog record data.

Summary

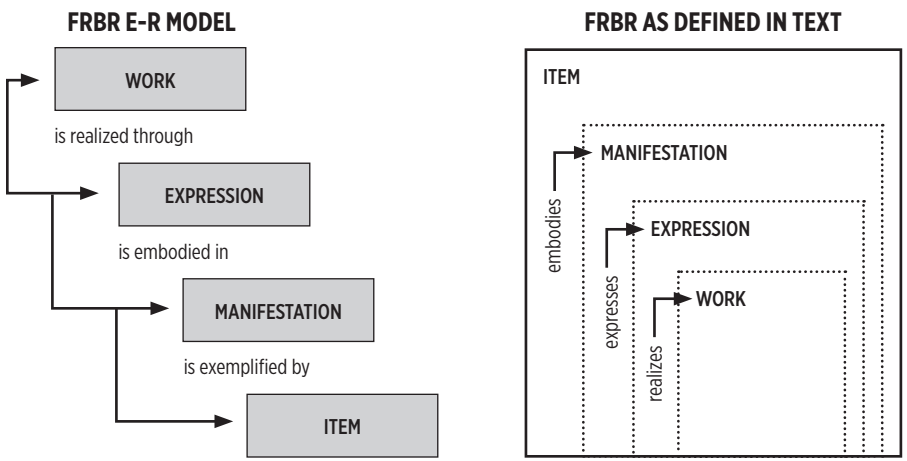
Group 1, as defined in the FRBR Final Report, exhibits some distinct differences between how it is defined in the text and how it is depicted in the entity-relation diagram. The entity-relation diagram shows four separate entities, and these entities are given each a separate set of attributes. This is consistent with a relational

database model that creates separate entities based on an analysis of redundancy of data elements. This does not mean that those entities are separate in "real life." We can use the example of a car model and color options. Obviously there is no car being offered that has no color, but color options can be stored in a separate table from car models in a database so that models and colors can be combined as needed in database output. This is similar to the separation of attributes into the WEMI entities in FRBR.

At the same time, the text of FRBR describes the entities as realizing or embodying or exemplifying other entities, which implies less separation than the entity-relation diagrams describe. Conceptually, an item is a concrete example of a manifestation that embodies the expression of a work. When you hold the item in your hand, you are holding something that has within it an entire Group 1. This statement results in a very different diagram from the entity-relation diagrams in the FRBR Final Report.

FIGURE 7.2

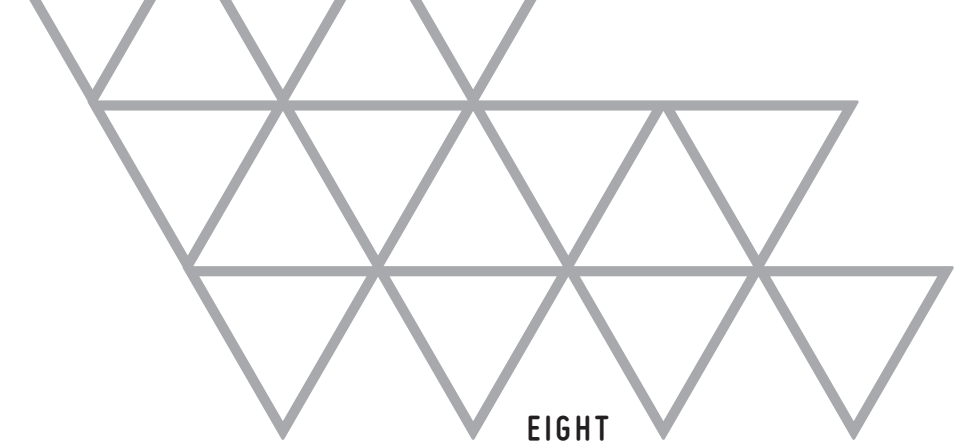
E-R diagrams versus definitions in the text



Both of these pictures of WEMI can be valid because they represent different views of the same thing. On the left of figure 7.2, we have a diagram of how WEMI will be expressed as stored metadata, in particular metadata that has been subjected to an entity-relation analysis. On the right, we have a diagram of a conceptual view of those same entities, without the need to separate the entities for the purposes of data efficiency. In fact, the diagram on the right, if it were

realized as a data format, would entail a great deal of redundancy of data. It is, by the way, very close to the data format that we use today. Which brings us back to FRBR as an entity-relation analysis of bibliographic data as it was created in libraries at the end of the twentieth century. If we add the two main authority files to the diagram on the right, we essentially have the equivalent of FRBR's entities prior to an entity-relation data analysis.

Looking at FRBR from this same point of view, it is clear that Group 1 represents a single thing, but that thing can be separated into parts for the convenience of managing nonredundant attributes. FRBR also describes the elements of a full bibliographic description. It is less clear, though, how those two views interact in practice, whether in the act of cataloging an item in hand, in sharing cataloging data, or in serving the library's users. Because FRBR was instantiated in the twenty-first-century cataloging rules, Resource Description and Access (RDA), the first of those questions may have been answered. The other questions, at the time of this writing, are still open.



EIGHT

DOES FRBR MEET FRBR'S OBJECTIVES?

> Many people want to evaluate a conceptual model like FRBR using "true or false" criteria. While one can say that a model is "true" to the extent that it explains accurately and "false" to the extent that it does not, this is not a very helpful way to look at these models. A more useful way to evaluate these models is to ask whether they are successful at fulfilling their purposes.
>
> —*Alison Carlyle, 2006, 266*

The FRBR model has resulted in unprecedented change in our thinking about catalogs and cataloging. Since the mid-nineteenth century, cataloging has focused on a unitary description of a physical item and a primary trinity of access points: title, author, and subject. This model has been disrupted by new technologies of communication, from the recording of sound and moving pictures to the digitization of nearly everything. Added to that disruption is the ever-increasing rate of production of creative resources of all types.

The previously stable view of the role of library catalogs began to be challenged in the latter half of the twentieth century as libraries struggled with these changes in their holdings. The key blow to the library's stability, however, was wrought by the Internet, the global network that allows instantaneous worldwide publication that defies the barriers of time and place. The slow conversation of books, which allows time for research, synthesis, creation, and production, is giving way to fast, short, mashed-up, and transitory personal and cultural expression.

Fortunately, some of the causes of this change are also potential factors in managing the change. Without computer technology libraries would not have been able to quickly add shared bibliographic records to their catalogs and thus to keep up—to the extent that they have—with the rate of resource production.

The FRBR model developed by the FRBR Study Group is that group's response to some of the late twentieth-century challenges. One can see in the conceptual model a strategy of "divide and conquer," a completely different approach from the "publication as unit of interest" of the cataloging concepts that guided the development of the AACR. In his essay in the collection *The Future of the Descriptive Cataloging Rules,* "AACR3? Not!" Michael Gorman, editor of AACR2, disputes the cataloging compromises proposed as a solution to the problem of multiple physical formats for the same content, saying that "Descriptions are of physical objects. . . . It is literally impossible to have a single description of two or more different physical objects" (Gorman 1998, 27).

FRBR is a direct challenge to the emphasis on a single unit defined by its physical description in descriptive cataloging rules. Although it doesn't ignore physical description, it does relegate it to a secondary role behind the placement of the content in a larger bibliographic context. FRBR introduces abstractions for works and expressions that had only appeared in earlier catalogs as collocation mechanisms, such as uniform titles.

As Carlyle says, though, the measure of success of a model is not its truth or falsity, but whether it achieves its goals. The objectives that were put forth by the Terms of Reference that guided the work of the FRBR Study Group give us one possible way to measure FRBR's success. This section looks at FRBR and its objectives, and asks:

- Does FRBR provide functional requirements?
- Does FRBR produce a national level bibliographic record?
- Does FRBR serve user needs?
- Does FRBR promote cataloging efficiency?
- Does FRBR aid data sharing?
- Is FRBR format neutral?

DOES FRBR PROVIDE FUNCTIONAL REQUIREMENTS?

> To begin with, we must note that our rules of descriptive cataloging contain no statement of the function which they are designed to serve. (Seymour Lubetzky, 1946, "Analysis of Current Descriptive Cataloging" 1946)

> The study has two primary objectives. The first is to provide a clearly defined, structured framework for relating the data that are recorded in bibliographic records to the needs of the users of those records. The second objective is to recommend a basic level of functionality for records created by national bibliographic agencies. (FRBR Final Report)

Seymour Lubetzky had two primary complaints about the cataloging rules and their outcome: first, that the rules stated what decisions catalogers should make, but not why the elements of the catalog were necessary; second, that the information in the catalog entry was not presented to the user in the order of importance, thus requiring the catalog user to wade through information of lesser importance in order to find those elements that were relevant. Both of these concerns speak to the role of the catalog in serving the needs of the user. Although Lubetzky did not articulate what those important elements were, it is easy to imagine that a user might find more value in an author added entry than the height of the book in centimeters.

One of the primary goals of FRBR was to rectify the lack of a functional justification by clearly aligning the data of the bibliographic record with the functions of the record and how it is used. This is the "functional requirements" of FRBR, and one cannot ignore that a stated purpose of the FRBR Study Group was to develop those functional requirements. It is also impossible to ignore the fact that functionality from the user perspective plays only a minor role in the final report.

Much is made of the user tasks find, identify, select, and obtain that are introduced in the FRBR Final Report. They are described there as "generic tasks" and are not made more specific. The description of them does not obviously motivate the selection of entities nor their attributes. In fact, the analysis of the user tasks is presented in the Final Report only after the entities and their attributes have been defined in some detail. It is significant that the document's structure is organized around the groups of entities, not the user tasks. The user tasks are covered in chapter 6 of the document, and are presented almost entirely as a comparison of the user tasks and the attributes that have been defined for the entities that are the main focus of the document. Attributes are rated as having a high level, medium level, or low level of importance for each of the broad user

tasks, as shown in figure 8.1. Some of the attributes are only of low importance, which brings into question why they are even included as necessary attributes. For example, the Work/Work relationships summarization, adaptation, transformation, and imitation are listed as low value for identify and select. Had the focus of the report truly been user needs, it is doubtful that those attributes would have been included.

FIGURE 8.1

A segment of a table of user tasks and FRBR attributes

	FIND				IDENTIFY			
	Work	Expression	Manifestation	Item	Work	Expression	Manifestation	Item
ATTRIBUTES OF MANIFESTATION								
Title of the manifestation	■	■	■		■	■	■	
Statement of responsibility	□	□	□		□	■	■	
Edition/issue designation						■	■	
Place of publication/distribution							○	
Publisher/distributor			○			□	■	
Date of publication/distribution	□	□	○		□	□	■	
Fabricator/manufacturer (note 1)			○				○	
Series of statement			□				■	
Form of carrier			□				■	
Extent of the carrier (note 2)							□	
Physical medium (note 3)							□	
Capture mode							○	
Dimensions of the carrier (note 4)							○	
Manifestation identifier			■			□	■	

The user tasks cover only the so-called "primary" entities, those of Group 1. There is no mention of how a user knows that she has found the correct entry for an author or a subject. In fact, both authors and subjects get short shrift throughout the FRBR Final Report, and it is presumed that they will be described more fully in the companion standards, Functional Requirements for Authority Data (FRAD) and Functional Requirements for Subject Authority Data (FRSAD).

FIGURE 8.1

A segment of a table of user tasks and FRBR attributes

	SELECT				OBTAIN			
	Work	Expression	Manifestation	Item	Work	Expression	Manifestation	Item
ATTRIBUTES OF MANIFESTATION								
Title of the manifestation	□	□	□				■	
Statement of responsibility	□	■	■				■	
Edition/issue designation		■	■				■	
Place of publication/distribution	○	○	○				■	
Publisher/distributor		○	□				■	
Date of publication/distribution	□	□	■				■	
Fabricator/manufacturer (note 1)			○					
Series of statement			□				■	
Form of carrier			■				■	
Extent of the carrier (note 2)		□	□					
Physical medium (note 3)			○					
Capture mode			□					
Dimensions of the carrier (note 4)			□				□	
Manifestation identifier							■	

Let's look at just a few aspects of the user task analysis. For the task find, the attributes of high value to find a work are:

- title of the work
- dependent component
- independent component
- persons/corporate bodies responsible for work
- entities treated as subject of work
- title of the manifestation

Attributes of high value to find a manifestation are:

- title of the manifestation
- manifestation identifier

Because no detailed explanation was given for these specific selections, one can only surmise what the FRBR Study Group was thinking. The title of the work, the "responsible parties," and the subjects are unsurprising, and mirror the ages-old goals of the catalog first expounded by Cutter. Less obvious are the dependent and independent components. These are essentially whole/part relationships, in which a dependent component would be a chapter in a book, whereas an independent component would be a monograph in a monographic series, or an article in a journal. These are real relationships, but because very few dependent parts are cataloged, the importance of these for the find task may be limited. As for the independent components, these are usually sought on their own, as in the case of journal articles. A direct relationship between the larger unit and the part is obviously a good idea, as primary elements for finding materials these aren't convincing.

It is also interesting that the manifestation title is an important element for finding the work, but the work title is not equally important for finding the manifestation. There is undoubtedly some logic behind that, but it is not explained.

For the manifestation, the "manifestation identifier" that is listed is most commonly the ISBN. This is an important data element, but I do wonder how often users (including library staff) approach the library catalog with an ISBN in hand (or head). The ISBN is, however, heavily utilized in automated processes, such as duplicate detection and retrieval of cover images from online sources. Because there is no definition of *users* in the document, it is not possible to know whom the group had in mind for the various data elements, nor can we know if some bibliographic attributes were specifically intended for automated processing.

The statement of responsibility is among the elements that have a moderate role for "find a manifestation." This is not a heading in library data, and I am confused by the assumptions the FRBR Study Group makes regarding the action of "finding." In fact, the report does not mention indexing, nor whether there is even an assumption that there are headings. Yet the find action does imply that some ability to search must exist, and the Final Report describes the elements of moderate value for find as those "typically used as a secondary search term." It isn't clear what "secondary search term" means, but presumably this is a term that can be used to limit results, as with the use of limiting elements in many catalogs by language, resource type, or other characteristic. The Study Group clearly harbored some implicit assumptions about system capabilities, but what these are is not made clear.

For manifestations, the only secondary element listed is "form of carrier." This is a typical secondary search term, but "form of expression" (i.e., whether it is text, music, or film) is not considered of any value for the find task, even though the advanced search of some systems includes the ability to limit by form.

Some of the moderate-value find elements are relationships: successors and supplements are included, but not adaptations or transformations. Most of these relationships are specific to serial publications. The FRBR-based catalog may not lead you from a parody to the thing parodied, but it should allow you to make the connection from a supplement to the thing it supplements. That serially published items need to be connected is rather obvious, but it isn't clear to me why these relationships are attributed to the find function and not, for example, identify or select.

The obtain task is not what I would have immediately expected. It would be logical to think that obtain refers only to items held or offered by the library, and that a primary element would have been the call number or some other identifier that leads directly to the item itself. Instead, the obtain task includes "acquire an entity through purchase, loan, etc." This wording is now included in the functions of the catalog listed in the 2009 IFLA International Cataloguing Principles. The previous version of the Principles, dated 1961, echoed Cutter's original functions of the catalog by limiting those functions to "whether the library contains a particular book." Thus the obtain task includes an expanded view of the role of the catalog. For this reason, obtain relates both to manifestations and items, and the key elements are those that define (or identify) the manifestation, including title, statement of responsibility, publisher, and series. For example, place of publication is considered of low value for identifying and selecting a manifestation, but is of high value for obtaining it. Oddly, place of publication is also of low value for selecting a work and an expression, even though, by the very definition of those

entities, no place of publication is possible. The user tasks might have benefited from an introductory discussion of the expected functions of the catalog. As it is, those functions must be teased out by reading between the lines.

The answer to the question "Does FRBR provide functional requirements?" is, for the most part, "no." The derivation of entities and attributes is part of the E-R analysis which dominates the FRBR Final Report, not of the analysis of user tasks. Everything about the document revolves around the E-R analysis, the entities, entity attributes, and the bibliographic relationships. When the user tasks are analyzed, after the FRBR model has been fully developed in earlier sections of the document, some of the attributes and relationships are shown to be of little importance. No functional requirements are given for the entities themselves. To have provided a truly functional analysis would have required making explicit some of the implicit assumptions that seemed to have been shared by the members of the FRBR Study Group. Examples of these are assumptions about system capabilities for search, and the goals of the catalog. The group obviously believed its own statement that FRBR was technology neutral, even though they used a specific technology, entity-relation analysis, to produce their results, and that they had to assume, but never explain, certain types of search capabilities that would make their attributes and relationships usable in some real implementation.

DOES FRBR PRODUCE A NATIONAL LEVEL BIBLIOGRAPHIC RECORD?

The FRBR Study Group was also tasked with determining the bibliographic elements of a basic level bibliographic record for national libraries (BLNB) based on the results of their analysis. It does so in chapter 7 of the group's final report. This chapter reads like an afterthought, however, because it isn't mentioned in the rest of the text, and oddly its content has very little to do with any other content in the report.

The BLNB defined in section 7.3 of the FRBR Final Report resembles ISBD more than it does FRBR. It uses the ISBD concept of "areas," such as "Title area" and "Edition area," which are not included in FRBR, and defines a short list of bibliographic elements for each area. The terms *work, expression,* and *manifestation* are not included in this section, nor does the BLNB make use of the attributes that are defined in FRBR. Where FRBR has "title of the manifestation," the BLNB has "title proper (including number/name of part)." The BLNB refers to "uniform titles," a term that is not used in FRBR, and is not defined anywhere in the report. Uniform title is also not included in ISBD, so it may not be known to those who would need to use the BLNB. The BLNB also includes a lengthy

"notes" area, even though no notes are included in the FRBR attributes. In fact, the BLNB appears to have very little relationship to FRBR at all.

The relationship between FRBR and ISBD is unclear. The FRBR report does cite ISBD as a source:

> The basic elements of the model developed for the study—the entities, attributes, and relationships—were derived from a logical analysis of the data that are typically reflected in bibliographic records. The basic elements of the model developed for the study—the entities, attributes, and relationships—were derived from a logical analysis of the data that are typically reflected in bibliographic records. The principal sources used in the analysis included the International Standard Bibliographic Descriptions (ISBDs). (FRBR Final Report, 4)

Although throughout the FRBR study ISBD is cited as a source of attributes for the entities, it is instructive to look briefly at a comparison of FRBR, ISBD, and the Basic Level National Bibliographic Record in chapter 7 of the FRBR Final Report. ISBD has ten areas and ninety-two attributes. FRBR has eighty-four attributes, plus thirty-four bibliographic relationships. The numbers alone do not provide a good comparison, however. FRBR does not include any note fields amongst its listed attributes. ISBD, on the other hand, lists thirty-two note types. The Basic Level National Bibliographic Record uses seven of the ISBD areas and adds four more areas for access points (names, titles, subjects, and series). It has fifty-three attributes, of which twenty-two are notes fields. Some of the notes describe bibliographic relationships that are included in FRBR. For example, the FRBR relationship "successor" in the BLNB becomes "note on edition and bibliographic history—successor."

The upshot is that the BLNB at least superficially resembles ISBD more than it does FRBR, which makes its appearance in the FRBR Final Report particularly jarring. It states that "the terms used to identify individual data elements correspond to the terms used in the *International Standard Bibliographic Descriptions* (ISBDs) and the *Guidelines for Authority and Reference Entries* (GARE)," but gives no explanation of why ISBD and not FRBR terms were used.

The introduction to chapter 7 states that the development of BLNB used bibliographic concepts from FRBR work, expression, and manifestation, but not item. In that section, the explanation for this decision is that it does not include item because this is a national library record. However, it isn't clear how that follows unless this analysis specifically does not cover archives or rare book materials, nor any other materials, like art and museum objects, for which item-level description is key.

It is worth noting that FRBR itself resembles a minimum bibliographic description. Compared to MARC 21, FRBR is clearly a high-level core of bibliographic data elements. FRBR has fewer than 100 attributes, while MARC 21 has nearly 1,500 unique data elements. Although FRBR defines a number of bibliographic relationships, like "successor to" or "adaptation of," there is only one relationship between a person and each FRBR Group 1 entity: creator (work), realizer (expression), producer (manifestation), and owner (item). In comparison, the list of relators that can be used for persons and corporate bodies in MARC 21 records has over 260 different terms.

The BLNB has not been instantiated by IFLA, even though this was a stated objective of the FRBR Study Group's work. It appears that this objective has not be fulfilled.

DOES FRBR SERVE USER NEEDS?

For a study that was purported to be user-centric, the user's absence is notable. There is no analysis of users; no mention of how varied the library user base is; no mention of children or elders or the disabled. Instead, to my mind, the FRBR Final Report reads as a study *by* catalogers *for* catalogers. Even more specifically, this is a study by catalogers at large research institutions. The list of participants in the FRBR Study Group does not include anyone from a public library, nor anyone from a non-Western country.

The FRBR document states:

> The study makes no *a priori* assumptions about the bibliographic record itself, either in terms of content or structure. It takes a user-focused approach to analyzing data requirements insofar as it endeavours to define in a systematic way what it is that the user expects to find information about in a bibliographic record and how that information is used. (FRBR Final Report, 3)

It cannot be coincidence that this study describes a bibliographic description that looks incredibly like the one that already existed at the time. As Le Boeuf says in the introduction to his 2005 book on FRBR: "FRBR models what we do, not what we should do." To say that the study began with a bibliographic blank slate, yet ended up describing the precise data elements that are in bibliographic standards like ISBD, defies credibility. The mere presence of elements like "statement of responsibility" shows that the needs of users did not drive the development of FRBR, because it's pretty clear that no nonlibrarian user would

call for a statement of responsibility. In fact, the FRBR Final Report states that "the attributes defined for the study were derived from a logical analysis of the data that are typically reflected in bibliographic records." This is circular reasoning, which should explain why some entities seem under-defined or underutilized. For example, there is a place entity for subject, but no place entity that can be used for place of publication, because the latter is not given authority control in current cataloging:

> Inasmuch as the model also defines *place* as an entity it would have been possible to define an additional relationship linking the entity *place* either directly to the *manifestation* or indirectly through the entities *person* and *corporate body* which in turn are linked through the production relationship to the *manifestation*. To produce a fully developed data model further definition of that kind would be appropriate. But for the purposes of this study it was deemed unnecessary to have the conceptual model reflect all such possibilities. (FRBR Final Report, 31)

How the FRBR Study Group addresses the question of what users need (or even what they want, which is not necessarily the same) is one that I find particularly disturbing. Not only were users not consulted about any aspect of this but the report provides no analysis of existing research.

So how did the FRBR Study Group study users? According to Olivia Madison:

> The FRBR Terms of Reference mandated that the study take into account the needs of a broad range of users and how they use bibliographic records. . . . One obvious option was to query, using a systematic methodology, a broad range of users, and draw conclusions from this analysis. Another option was to use our collective knowledge of the various types of users from the working group membership and commentators, as well as to draw upon experts in the fields to provide necessary user perspectives and conclusions. The Study Group decided in favor of the latter approach. (Madison 2005, 28–29)

My note to myself on that page reads: "We don't need no stinkin' user surveys." The FRBR Final Report implied that some study was done of the literature:

> The assessment of importance of each attribute or relationship to a given user task that is reflected in the tables was based in large part on the knowledge and experience of the study group members and consultants, supplemented by evidence in the library science literature gathered from empirical research, as

> well as by assessments made by several experts outside the study group. (FRBR Final Report, 83)

In other words, the study of user needs was done without studying users. But in fact the Terms of Reference that gave the group its charge also says very little about users and does not require the group to study them. In addition, the development of the user tasks was not dictated by the Terms of Reference.

The document itself is not structured around the user tasks, and those tasks are defined only at a very high level before the document begins its description of the three groups of entities of bibliographic control. Within the document sections that describe the entities, the attributes, and the relationships between entities, there is no discussion of how the user tasks inform these aspects of the model. Although there is a section devoted to the relationship between the user tasks and the attributes, there is no discussion of *how* different attributes help users make their decisions, only a statement that they *do* provide information for users. For example, in section 6 on user tasks, the identify task is described in this way:

> 1. the attribute by definition serves to identify the entity (e.g., *manifestation* identifier, *item* identifier);
> 2. the attribute or relationship forms part of the minimal set of attributes and relationships that for the majority of cases will serve, in the absence of a unique identifier, to differentiate entities that have a number of common characteristics (e.g., the minimal set of attributes sufficient to differentiate one *manifestation* from another in the majority of cases comprises title, statement of responsibility, edition/issue designation, publisher/distributor, date of publication/distribution, series statement, and form of carrier). (FRBR Final Report, 85)

This leads one to wonder just how the user is defined. The identifiers in point one are generally for machine processing, not for humans, although occasionally a human may come to the catalog in possession of a relevant identifier, like an ISBN. Are machines included in the group's concept of users? On point two, unless the user is a trained cataloger, this is a completely unfounded set of assumptions. How do users *actually* identify manifestations? Do nonlibrarian users pay any attention to the statement of responsibility? Do they understand the series statement? In fact, do they even see these data elements while making their choice? Many systems only display these in a full display, and users must make their choice from the results list with brief bibliographic displays.

The main differences between FRBR and our current bibliographic records are the division of the bibliographic description into four entities, and the potential

for the creation of linking relationships between bibliographic entities. Although the latter holds promise, there does not yet exist a system that would provide evidence of user interaction with these relationships. There have been some studies providing system displays that use the Group 1 entities to cluster retrievals, and these have shown that the presentation of clusters of works is often preferred to the list of redundant editions that catalogs generally provide. This begs the question, however, of whether the FRBR Group 1 coincides with the user view of the bibliographic world.

Two studies by Jan Pisanski and Maja Žumer (2010) on residents of the Slovenian capital of Ljubljana investigated this question of whether the FRBR Study Group's view of Group 1 entities is consistent with how users view the bibliographic universe. They titled their studies "Mental Models of the Bibliographic Universe." The studies used card sets that presented library users with bibliographic information consistent with the entities of Group 1. For example, using Dan Brown's *Da Vinci Code,* there were cards for:

- Dan Brown's Da Vinci Code
- Dan Brown's Da Vinci Code, published by Doubleday in New York in 2003
- The Slovenian translation of The Da Vinci Code by Nataša Müller
- The movie of The Da Vinci Code starring Tom Hanks, in English with Slovenian subtitles

and others. Note that the users were not given individual Group 1 entities, but instead were presented with either a work; a work and an expression; a work, expression, and manifestation; or the entire Group 1 from work to item. This is consistent with the conceptual view of Group 1, although not its entity-relation view.

The users were asked to place these cards in order from the most abstract to the most concrete. This instruction was difficult for many users because they hadn't previously applied the concepts of "abstract" and "concrete" to bibliographic items.

In a second study, users were asked whether, of a pair of items, either item was a reasonable substitute for the other.

The authors made some key discoveries:

1. Users have many different views of the bibliographic universe, but where users had a common view, it generally was a FRBR-like view of the "progression" from a general concept (work) to individual publications (manifestation) and lastly to specific items (such as signed copies).

2. What mattered most to users was the language of the text, the form (book versus film), and the contents (illustrations versus no illustrations).
3. Users have a strong sense of "original work" which was the item that they placed at the "top" of the hierarchy, and for them took the place of the FRBR abstraction, work.
4. Users seek items at the FRBR level of expression for the most part, and consider the related manifestations to be acceptable substitutes for each other.

These results confirm some of the assumptions built into FRBR and into bibliographic cataloging rules, but they also reveal interesting variations. First, it should not be surprising that the results of the card sorting exercise did not turn out a single view of the tested items. Within the cataloging community there is debate over the exact definitions of work and expression (less so of manifestation), in spite of there being a shared knowledge of current cataloging concepts within that community. Nonlibrarian users (and one could extend that to librarians who have not been trained as catalogers) should be expected to bring a wide range of views to the table. The test group, though, was given strong hints through the experiment, which represented cataloging that had already surfaced a particular view of the bibliographic items based on FRBR's Group 1.

The discovery that people have a sense of the "original work" is quite sensible. There is nothing in FRBR, nor in the RDA cataloging rules based on FRBR, which directly addresses the question of "original work." Yet there is obviously a progression from the creative form that comes directly off the efforts of the inventor of the work and all versions that follow it. For Patrick Wilson's concept of bibliographic families, the original work would be the family's progenitor. In the case of translations of texts, there is an implicit concept of an original form that is translated to some other language. This progression is less strong in other areas, such as in music, where there may not be a recording that one could think of as an original performance of the piece. The separation between film and text can be clear, as in a film version of *Da Vinci Code,* but it is less clear for the *Star Trek* series of books that were written after the television series aired.

The conclusion that users seek language materials at the expression level is quite logical, and this coincides with Shoici Taniguchi's theory of an expression-dominant model for texts. The work is an abstraction without expression, and in practice the work would represent all language versions of an oft-translated work, or all editions of a work that had been updated, such as a yearly almanac. Most users have a language preference for reading, and therefore the work set

that includes all languages of the text would not be useful. In the case of different revised editions, such as with textbooks or reference books (which was not tested for in the Pisanski and Žumer studies), it would be unusual for a user to be seeking all editions, and even more unusual for a user to be seeking something other than the most recent edition available.

For the user at the catalog, the FRBR user task find may not visibly change if FRBR is adopted in the future. First, we must assume that users will not be aware of the Group 1 structure, but will search as they do today, which means either by keyword, which pools words from all of the searchable fields in the record, or by creator or contributor, title, or subject.

I am assuming that these searches will continue to work in this way, with the user not being required to know what Group 1 entity the search should go against. This would be consistent with how searching is done in catalogs that are not organized around the FRBR entities. For example, in most online catalogs, the title index includes all or most title elements from the record; therefore the user searches the work title and the manifestation title with one search along with any alternate titles. The same is true of the various creators and contributors to the resource, who are often searched together in the same index, whether authors, illustrators, composers, or librettists. This practice will probably remain, even though in FRBR primary creators are linked to the FRBR work and some secondary creators such as translators and illustrators, are linked to the expression or even the manifestation.

Where changes are anticipated in the catalog, however, are in the user displays. There is a general assumption that users will not see a manifestation-level display as they do today, but will be given a view that takes advantage of FRBR Group 1 as a way to gather all versions of the work together in a new kind of collocation using the primary author and the work title. However, there may need to be decisions based on the language of the catalog or of the user. For example, the work title Война и миръ (*War and Peace*) may not be useful in a catalog aimed at English-language speakers, yet that is the correct work title as defined in the cataloging rules. Whether that title should be displayed to the user or not is a question for user-interface designers. The expression entity in catalogs today is often represented by a facet that allows a user to narrow the retrieved set by language. Because FRBR is only a conceptual model, how the FRBR entities might be used in future catalogs is not specified, and the document does not provide any suggested display forms.

For some materials, the FRBR user view may be a vast improvement over the record-level view of today's online catalogs. The "Scherzo" project at Indiana

University, as reported by Handesty et al. (2012), developed a FRBRized catalog of music materials and did comparative user testing between the FRBRized and the traditional catalog. Although some of the results were mixed, they concluded that users preferred the FRBRized catalog.

The benefits of an FRBRized display are often used as an argument for moving to FRBR as the basis for cataloging. In fact, the FRBRized displays that have been tested have mainly been built from bibliographic data that was not cataloged according to FRBR principles or stored digitally as FRBR entities. WorldCat is developing a work-based display from its pre-FRBR bibliographic records, as did the library vendor VTLS. Outside of the library world, bibliographic databases like LibraryThing and OpenLibrary have developed work-based displays from data that was created without the FRBR entities as concepts. A FRBRized display that used data that was originally designed around FRBR principles may be superior, but there may be ways to take advantage of some FRBR concepts without having to entirely re-engineer library systems.

There is one key user-related concept that gets no mention in the FRBR Final Report, and that is "collocation." Today's catalog record is a single metadata container that carries all of the entities described by the FRBR Final Report. The primary focus of this record is what FRBR would consider to be a manifestation, and in most systems the dataset returned from a search is a group of manifestation-focused records presented in some order (ranked, by date, or alphabetical). With our bibliographic universe organized as defined in FRBR Group 1, however, there are new options relating both to retrieval and to display. FRBR has the potential to restore the context of bibliographic relationships that existed in the collocation function of the card catalog. This loss of collocation has been a source of frustration for catalogers, as I have learned from many conversations about library systems. Martha Yee described online catalogs as mere "finding lists," and suggested that "FRBRization" could produce a list of works, expressions, and manifestations related to a user's search so that users could make use of that bibliographic context to select resources of interest (Yee 2005).

Perhaps it is deemed to be obvious from the inclusion of bibliographic relationships that a major function of FRBR is to reintroduce the collocation function that was lost to the online catalog, when discovery through the linear order of the physical catalog was replaced by database-based retrieval of individual bibliographic records. Collocation has always been a surfacing of relationships between bibliographic items, although in the physical catalog the relationships could not be made explicit. It was left to the user to intuit the meaning of the physical proximity between cards, if there was one at all. The separation of bibliographic data into

entities is emphasized in the FRBR Final Report, although the real potential for user service comes from the ability to not only collocate related bibliographic items, but to make explicit what that relationship is. Although the Pisanski and Žumer studies look mainly at how users organize the elements of FRBR Group 1, they show, perhaps inadvertently, that users are cognizant of bibliographic relationships like translations and adaptations, even though they may not have a conscious awareness of those relationships and what they might mean.

The cataloging rules based on FRBR attempt to create a shared concept of the entities and how they are described so that libraries can share equivalent bibliographic data. What the Pisanski and Žumer studies tell us is that we cannot assume that untrained users will immediately understand how libraries have divided up the bibliographic universe because it may differ from their own underlying assumptions, although with the language materials that they studied there is much overlap between the user view and the FRBR Group 1 concepts. The advantage of consistency is that it provides a platform that can be learned; whether the general public will benefit from that, and who within that general public will benefit, is not known.

DOES FRBR PROMOTE CATALOGING EFFICIENCY?

One of the primary reasons behind the development of FRBR was to create a more efficient environment for cataloging. This came out of the 1990 Stockholm IFLA meeting. The combination of lowered budgets and increased information resource output creates a perfect storm for libraries, and especially their technical services departments. So it makes sense to ask whether FRBR is likely to increase cataloger efficiency. This is one of the goals behind the FRBR Group 1, because it makes it possible to share cataloging at the work and expression levels, and not to repeat data elements for works and expressions for each manifestation. In a system with separately stored works and expressions that are linked to manifestations, the cataloger would not have to repeat the work and expression information, but only provide a link from the manifestation to the expression. This is seen as a savings of time for catalogers, as well as potentially a point of efficiency for systems. Although some libraries have been cataloging under the RDA cataloging rules for over two years now, existing systems do not take advantage of the entity-relation structure of FRBR, so they have not gained experience with the ability to share separate descriptions of works and expressions.

One of the difficulties in answering this question is that libraries are not a homogeneous group, and cataloging is not the same activity throughout the

library world. In particular, national libraries and large research libraries probably do the greatest share of original cataloging, and the most detailed cataloging. A considerable number of smaller libraries only copy catalog data created by others for common published materials. Others purchase a great deal of their cataloging from the library vendors who supply their books and other materials. Within the libraries that rely heavily on copy cataloging, there is variation in how much editing is done of the cataloging data to customize it for the individual library.

How much one gains from the division of bibliographic data into entities depends on how often one can reuse entities that already exist. That is not easy to assess. More than one study has been done to try to quantify the proportion of works that are in defined bibliographic relationships with other works. Richard Smiraglia (2001) studied a small number of general academic libraries as well as libraries specializing in music and theology. Not surprisingly, the specialist libraries showed different patterns of bibliographic family membership. Sherry Vellucci's 1997 study of a music library showed that over 85 percent of the holdings were in a bibliographic family that could involve some sharing of data. In a general academic library the percentage may be as much as 50 percent. In OCLC's WorldCat database, which is more varied than most catalogs due to its international nature, about 30 percent of manifestations are in a bibliographic family, as per the study done by Brian Lavoie in 2001, as well as more recent statistics comparing numbers of works with total records.

Smiraglia also shows that having family members is a function of time: older works are more likely to be in families. A rather obvious, but worthwhile, point to prove is that that works go out of print and are republished and repurchased; and that derivative works occur only after time has shown that the work is worthy of further study and derivation. In the theology library collections studied by Smiraglia the mean age of works was 114 and 125 years, while general collections had a mean age of 32 years. Collections in non-research libraries are probably much younger than those studied by Smiraglia. Smiraglia's studies also showed that form and genre were not predictors of family relationships in the libraries studied.

One can conclude that the nature of the material being collected has a great influence on the incidence and size of bibliographic families. None of the studies, however, looked at nonacademic libraries, whose collections are not analogous to academic libraries, nor did they analyze topic areas other than theology and music. With a reminder that these are studies of library holdings and not of publication patterns, one could expect that where scientific works are updated and republished, older editions might not remain on the shelves, and therefore are not included in the catalog. Small and medium public libraries may be

expected to have fewer family members due to the exigencies of limited shelf space and the mission to serve the reading, listening, and viewing public more than research activities.

In the previously mentioned analysis of manifestations and works in WorldCat, published by Brian Lavoie and colleagues in 2001, yielded this information:

Number of manifestations (records)	46,767,913
Number of works	32,000,000
Average number of works per manifestation	1.5
Number of works with only a single manifestation	78%
Number of works with seven manifestations or less	99%
Number of works with more than twenty manifestations	1%

Because of the nature of the WorldCat database, which represents the holdings of many tens of thousands of libraries worldwide, these figures cannot be extrapolated to be meaningful for a single library. On the other hand, because of its breadth, WorldCat gives us a reasonable picture of publishing patterns over the combined universe of library holdings. This tells us that a large number of published items appear in only one version, and therefore the manifestation-expression-work ratio for those items is 1:1:1. Using the 2001 OCLC figures, 78 percent of the items in WorldCat are single manifestations of this type. Only 1 percent of the identified works have been produced in a large number of manifestations. The key question now becomes: What do these figures mean for the cataloger workflow, and would a cataloging system that makes work and expression data available for reuse result in greater efficiencies than we have with a cataloging model that is manifestation-centric?

To answer this question we would have to know how many previously described works a cataloger is likely to encounter when processing new items today. The figures that we have only reflect the cumulated past, not the present or future publishing output that catalogers confront. Regardless of the number of expressions and works in a bibliographic family, every new manifestation must be described. We should be able to conclude that there is no change in terms of cataloging efficiency in the description of manifestations. The new efficiencies would be found in the aspects of the bibliographic record that describe the expression and the work. There are three possible situations that a cataloger doing original cataloging may encounter:

1. *A manifestation that represents a new expression and a new work*. This would presumably be the case for newly published first editions. The cataloger creates the necessary data elements for the manifestation, the expression, and the work.
2. *A manifestation that represents a new expression of an existing work*. The cataloger creates the necessary data elements for the manifestation, the new expression, and links to the existing work.
3. *A manifestation that represents an existing expression (and thus its related work)*.The cataloger creates the data describing the manifestation, and links the manifestation to the existing expression.

All of these represent the activities of original cataloging. It does not appear that the activity of a cataloger who copies catalog data created by others is greatly changed from what is done today, which is to locate an existing bibliographic record for the manifestation in hand and to utilize that data for the local catalog.

Whether or not FRBR is efficient for catalogers, and whether the existence of FRBR work data saves time, depends entirely on how often catalogers encounter situations two and three, above. The WorldCat study only yields statistics for manifestations and works, not expressions, and that study presents a snapshot of the WorldCat database, which may not represent the bibliographic situations encountered on a daily basis by today's cataloger. Obviously, the majority of older published works have already been described, so what interests us is how often those are reissued as new manifestations, and whether today's publishing patterns are producing more or fewer instances of reuse of expressions and works. In addition, catalogers in different types of libraries (for example, medium-sized public libraries versus large research libraries) will encounter a different pattern of publication types. None of the above can be applied to catalogers of unique archival materials that do not benefit from the sharing of bibliographic description. The gain in cataloger efficiency may be greater for some materials, like music recordings, which have a high degree of repetition of works.

It is necessary here to reflect on the effect of the changing nature of library materials from solely physical to digital. One of the motivations for reviewing library cataloging, beginning in the early 1990s, was the recognition that libraries now often receive the same content in more than one physical format. This was prescient at the time because only a few digital formats were commonly available. Today, with the increase in the number of e-books and e-journals, it is quite common for a library to offer both a physical and a digital copy of the same resource, and even multiple forms of the digital version. The cataloging

rules regard these as different manifestations, but some libraries have been loath to confront users with multiple records for the same content in different forms. This was the problem called "multiple versions" that was much discussed in the 1990s. Many libraries have "solved" this problem by providing links to digital versions on the records for the hard copy version, thus providing the user with a single point of entry for either format. This violates the cataloging rules, however, because a manifestation is defined by its physical format. Using FRBR Group 1 as separate described entities means that all of the multiple physical versions can be associated with a single expression that is subordinate to a single work. This could allow grouping of multiple versions under work or expression in library catalogs, while still adhering to the emphasis on physical formats that motivates the cataloging rules. Of course, there could be many ways to achieve this kind of grouping; the FRBR entity-relation structure is only one of those.

We can conclude that we need much more data before we can answer the question of whether FRBR Group 1 actually saves time for some catalogers. In addition, whether the entire FRBR entity-relation model results in greater efficiencies may depend more on systems designs than on anything specific in the FRBR model. In current cataloging, name and subject authorities are already analogous to FRBR's Groups 2 and 3, and these are shared as records representing specific members of those entity types. While indeed the FRBR entity-relation model appears to promote efficiency in cataloging, there is as yet no concrete evidence that it will.

DOES FRBR AID DATA SHARING?

Another question, however, is how well FRBR works for sharing of bibliographic data between library systems. In the early days of modern cataloging, beginning in the middle of the nineteenth century, each library was an entirely separate entity, and the bibliographic universe was limited to the holdings of that library. Today library data is shared internationally.

There is a tension between the shared bibliographic universe and navigation within the catalog of a single library. This conflict could become even greater with the application of FRBR relationships as machine-actionable links. In today's bibliographic records, many of the relationships that we assume will be links in FRBR are presented as notes ("Translation from the German." "Based on the book by . . ."") or as assumed from edition statements ("3rd ed., revised and enlarged"). These notes do not provide useful linking, but they do provide information for the user even when the library does not hold the item referred to in the note.

Links alone will not provide that information for catalog users in cases where the links do not resolve to another bibliographic description in the library's catalog. Ultimately, users need both information and direct links where available. The FRBR entity-relation model does not address implementation issues of this nature, so we should assume that FRBR is a starting point upon which development may take place, even if that development results in a modified model.

The FRBR relationships are universally bibliographic in nature; they definitely go beyond the inventory role of the library catalog. In a large research library the holdings may be extensive enough that the disconnect between bibliography and inventory remains at tolerable levels. Even so, what is the library to do with machine-actionable relationships to items it does not own?

There is an ideal bibliographic universe where every work ever produced is placed in its proper place, and the whole can be navigated intelligently. There is no real version of that bibliographic universe, however. Nonetheless there is a desire to use today's technology to create the "library of everything," though "everything" has a lot of ragged edges. It also has a lot of individual libraries with different populations. FRBR as written in the Final Report appears to address an abstract bibliographic universe, which may or may not coincide with the reality of individual libraries. The tension between bibliography and the catalog will have to be resolved. FRBR operates at an abstract level that needs to be reinterpreted for the many thousands of use cases that exist in libraries, and still allow them to share cataloging copy.

This tells me that we need to define the relationship between the library catalog and the bibliographic universe. I can imagine a multitiered design that can move a user from the inventory of the single library to related bibliographic items that are nearby or otherwise available (e.g., for purchase). I can also imagine a linked bibliographic space that is not dependent on library holdings, but that is about the relationships between bibliographic items. This would need to go beyond FRBR, because some of the more useful relationships (e.g., "cites" or "quotes") are not included there. Connecting this "cloud" of bibliographic relationships to actual availability in specific libraries or bookstores would be essential. Whether we can develop the technology to realize this vision remains to be seen.

IS FRBR FORMAT NEUTRAL?

A model like FRBR needs to accommodate all of the types of resources that will be included in the library catalog. This is a difficult requirement because libraries and archives manage a wide variety of materials whose bibliographic

natures are far from uniform. Although the FRBR Study Group stated that FRBR was designed to describe a model that could be applied to any type of bibliographic data, numerous articles have discussed how the conceptual model would need to be reinterpreted for specific resources types. Some of those types are presented below.

Serials cataloging deals with a particularly complex set of relationships: the whole/part relationships between the resources that users generally seek, the articles, and the continuing resource that is the serial itself. There is also the very difficult question of what is the serial work, and when does a serial become a new serial?

There are relationships defined in FRBR that are suited to serial publications, such as Successor and Supplement, and properties such as Sequencing Pattern and Expected Regularity of Issue. As is often the case, however, the difficulties come in the interpretation of the Group 1 entities.

Paul Weiss and Steve Shadle (2007) look at "FRBR in the Real World." They point out that it isn't obvious that a serial is itself a work in the FRBR sense of that term, that is, "a distinct intellectual or artistic creation." It is easy to see serials as aggregations of works, but the serial itself appears to lack that distinctness that FRBR requires of works. As we'll see in a later section, aggregations of works present a number of problems in FRBR, including the fact that each aggregated work itself is defined as having the full complement of Group 1 entities, which means that whole/part relationships between the larger entity, the serial, and the part entities, the articles, can be quite complex. This complexity is only the beginning, because serials publishing provides some stunningly difficult examples, such as serials within serials, supplements to serials, and a wide variety of enumeration patterns.

Kristin Antelman (2004) also addresses the question of seriality and the work. She notes that the main identifiers for works, both in FRBR and in earlier cataloging theory, such as author and title, are not strong identifiers for serials. Serials often do not have "authors" per se, and because their titles so often change, one journal of serials cataloging is itself titled *Title Varies*. Serials also have been the main resource type to be provided in digital format as well as print, with the result that libraries often have both print and digital subscriptions to the same content but packaged differently. Antelman recommends "asserting bibliographic control over a higher level of abstraction than has been our practice," and putting

a "greater emphasis on relationships between abstract entities and less on the identification of the physical item."

Ed Jones, in his article on "The FRBR Model as Applied to Continuing Resources" (2005), points out that the boundaries between work, expression, and manifestation in continuing resources differ from that of other materials because changes in key elements, such as the title, can take place in one expression or one manifestation but not in the others. Because serials cataloging leans heavily on titles for the identity of serials, this poses a conflict between the title-based definition of work in serials cataloging and the definition of work in FRBR. Jones's argument is not that the bibliographic entities in FRBR are problematic, but that they are not in accord with the rules for serials cataloging. Either FRBR or the serials definition of the work as being title-based would need to change. This sounds radical, but it should not be surprising that a new conceptual model could lead to changes in cataloging practice, or that cataloging practice should inform modifications to a conceptual model.

One possible positive outcome of FRBR or a FRBR-like linking model is that it may be easier to link articles and the journals they are published in. Laura Krier (2012) sees advantages in the use of linked data for serials. Krier's analysis isn't specific to FRBR, but FRBR relationships are one possible way that serials linked data could be achieved.

Music is an interesting cataloging case even as presented in the FRBR Final Report itself. The relationship of work and expression in music often takes place as a performance, and individual works can have a very large number of expressions. Yet, music cataloging (primarily for classical music) makes more use of work titles than any other specialized type, which means that the concept of the work is strong among librarians and probably also among users.

Raymond Schmidt, however, takes a look at jazz and improvisation in his "Composing in Real Time: Jazz Performances as 'Works' in the FRBR Model" (2012). If each performance of a jazz musical piece brings in new sounds, the difference between a work and an expression is blurred, and there are then questions of whether or when a performance creates a new work. Musicians and listeners recognize familiar tunes, and might consider two performances to be of the same work. Yet the musicians who improvise performances or engage in sampling of the works of others often consider themselves to be creating new works even though a recognized "air" is present.

The Variations Digital Music Library 2 Project at Indiana University, as reported by Jenn Riley at the 2008 International Society for Music Information Retrieval conference, analyzed the FRBR attributes for each WEMI entity and had a number of modifications that were required to express the music materials. That analysis concluded that the FRBR concept of a work may be suitable for the "canon of Western art music, where composition takes precedence over any given performance of it." However, they found that there were additions needed to FRBR work for jazz, World, and non-Western music, which when music is not in written form are the musical equivalent of an oral tradition. To describe these musical forms correctly, the FRBR work would need the additional attributes of language and place of composition. They also suggested that musical key was needed for both works and expressions. (In FRBR it is only an attribute of work.)

Music publishing also presents some difficulties. Many works are issued in a kind of "bound with" relationship with other works. Music manifestations have the added complication of accompanying materials, such as liner notes, which need a language designation in the manifestation, but not in the expression. Riley provided this comparison of FRBR and the Variations project's versions two and three:

VARIATIONS 2/3 ENTITY	FRBR GROUP 1 ENTITY
Work (more concrete than FRBR Work)	Work
Instantiation (can only appear on one Container)	Expression
Container (includes some copy-specific data)	Manifestation
Media Object (defined as a digital file)	Item

This illustrates that the FRBR work may be an organizing concept rather than a fixed "thing" that organizes at a different level of abstraction within different contexts. In classical music, a performance based on the same score could be considered the same work, while in jazz or hip-hop a performance might indeed be the creation of a new work.

Martha Yee, a cataloger at the University of California at Los Angeles film archives, found that the FRBR definition of work accords with that of the moving images cataloging community, which is that the translation of a textual or musical work to film always creates a new work. However, this view illustrates a distinct

difference between film catalogers and other catalogers—most notability those working with music materials:

> Film catalogers consider a film of a performed work to be a new work related to the previously existing text for the performed work. Thus a film cataloger considers Bergman's *Magic Flute* to be a new film work, not an edition of Mozart's opera. Music catalogers disagree; to them, Bergman's film is still primarily Mozart's work; that is, not a new work, but a new expression of Mozart's work. (Yee 2007, 124)

As Yee describes FRBR, "Change from any other GMD [General Material Designation or physical type] . . . into the moving image GMD . . . creates a new work by [the] FRBR definition" (Yee 2007, 121). This contrasts with the music cataloging view that a new performance or recording is a new expression. In a library catalog containing both music and film materials, these two views cannot coexist.

James M. Turner and Abby A. Goodrum (2009) address the many types of editing that can be performed on moving pictures, including removing content for reasons of audience (profanity, nudity) or to make the film fit the time allotted. These, according to the authors, are treated as new manifestations, where in the text world these would undoubtedly be considered new expressions. At the same time, there is interest in what Turner and Goodrum call "frame-level" access to film, with examples like the Zapruder film of the Kennedy assassination and various films of the 9/11 events. These are spontaneous creations, sometimes even accidental, and so the designation of a creative work component may be difficult. However, should such unplanned footage be included in a documentary, its nature appears to change. Three people standing side-by-side at a scene may take nearly identical photos or films of an event. How many works are there? There are also the multitude of similar scenes shot for a commercial film which may or may not be included in the film, and yet may be reconstructed later into a new version of the film as the "director's cut."

Although library standards attempt to cover a wide range of library and museum materials, there is a significant distance between the description of published materials and the treatment of unique materials, such as works of art. The cultural object community has its own cataloging rules, *Cataloging Cultural Objects* (CCO), which differs considerably from the rules for printed and published materials.

In an analysis of FRBR in relation to art materials, Baca and Clarke (2007) point out that the separation of work, expression, manifestation, and item is not applicable for many of the materials they catalog. The CCO rules define a work as a human-made object, meant to be perceived through the sense of sight. They go on to say

> Because of this sense of physicality, *work, expression, manifestation and item* are one and the same for many cultural works, which are embodied in a *single material object* and not in an *abstract entity.* (Baca, 104; emphasis in the original)

The CCO entity diagram has a central work, with authority-like entities for names, geographical entities, concepts, and subjects. The central work entity encompasses all of what in FRBR makes up Group 1.

The authors concede that FRBR Group 1 entities may be suitable at times to conceptual art and artwork that takes place over time and in different forms, where the artist's concept and the execution of the art are not a single action. These do not always break down into the WEMI categories of FRBR, however. It appears that the description of artworks may need to make use of a variable interpretation of the Group 1 entities, as necessitated by different types of artworks. A simple example is in the role of color in artworks. In FRBR, color ("colour") is an attribute of the manifestation, and is defined as "used in the production of an image" (FRBR Final Report, 46). For some images, such as abstract paintings, color is the content of the work and defines the work. One need only look at the paintings that are variations on the theme of "white on white" to understand that color can be the idea of the work.

The CCO view has equivalents to Group 2 entities, but it differs in how it treats what in FRBR are the Group 3 subject entities. CCO uses two subject entities, geographic place and concept, but does not list object or event as primary entities. The subject authority in CCO is used to depict the subject matter of the artwork, that is, what it is about. The concept authority is a thesaurus of terms used in description, perhaps similar to the various controlled descriptive terms developed for MARC records or for the RDA cataloging code. This thesaurus includes terms for objects, materials, activities, agents, physical attributes, and time periods. In the final analysis, much of the same descriptive territory is covered although the organization of the elements is somewhat different.

Because maps are often published in editions, the use of the FRBR expression entity promises to help organize these materials. Two articles that focus on maps and FRBR, one by Ruth Kalf and the other by Scott R. McEathron, illustrate the use of FRBR to describe maps of the sixteenth through eighteenth centuries. This was a time when maps were uncommon and the few that existed were copied and reproduced in different languages. Those maps coincide relatively well with the FRBR document-oriented model. For modern machine-rendered maps, which can be produced easily in different sizes and scales, Kalf questions whether scale, which seems appropriate at a FRBR expression level for the older maps, could not be described at a manifestation level. This exemplifies the need for flexible definitions of the FRBR entities even within a single cataloging specialty. In this case, reproduction technology has changed how the resource is created and realized, and a change in map scale is now more closely analogous to converting a document from one word processing format to another.

Summary

It is quite possible that few of those among us who consider themselves to be reasonably familiar with FRBR are actually aware of the original goals of the project as stated in the Terms of Reference, even though these are mentioned in the FRBR Final Report. It is also possible that many are not even aware of the seventh section of the Final Report that defines a minimal level record, which was the primary motivation for the creation of the FRBR Study Group.

The FRBR Final Report provides a revolutionary view of bibliographic data as conceived in the service of the library catalog. It will probably be viewed by future generations of librarians as a watershed moment in the history of the library catalog. The gap, however, between the original goals and the FRBR Final Report is great. The FRBR Study Group went beyond the originally stated goals in an attempt to solve problems that had been bubbling to the surface of cataloging practice over at least the previous half century. The task that arose from the 1990 Stockholm meeting became the precipitating factor for a general review of descriptive cataloging and its purpose.

That the FRBR Final Report overstepped the original goals is of historical interest, and should help us understand that the context for FRBR is not in a single meeting in 1990 but in all that came before it.

NINE

SOME ISSUES THAT ARISE IN FRBR

There are some fundamental problem areas in FRBR as I read the document. There have been hints some of these in the preceding description of the FRBR entities: marked differences between the underlying structure of the three groups and a lack of cohesion within the groups themselves (e.g., there are no shared qualities among group members). Some of these issues just don't feel quite right; others pose problems in implementation. Among the truly difficult problems are those of inheritance and hierarchy, of the disjoint nature of the Group 1 entities, and how bibliographic aggregates fit (or do not fit) into the FRBR model.

INHERITANCE AND HIERARCHY

It is a common assumption that the four entities in Group 1 represent a hierarchy with inheritance, even among FRBR Study Group members: "Expressions inherit properties from works and manifestations inherit properties from expressions, not vice versa" (O'Neill et al., "Final Report on the Working Group on

Aggregates"). In fact, it might be best if they did, but they don't, as Renear and Choi demonstrated in their paper "Modeling Our Understanding, Understanding Our Models" (2006). In fact, the E-R modeling technique used in FRBR *cannot* express inheritance. The four entities of Group 1 are linked through relationships in a kind of bibliographic daisy chain; therefore, for example, a FRBR manifestation has access to the attributes of a FRBR work through the link to the FRBR expression. But as described, a manifestation cannot "have" an author or subjects because those are only attributes of a work.

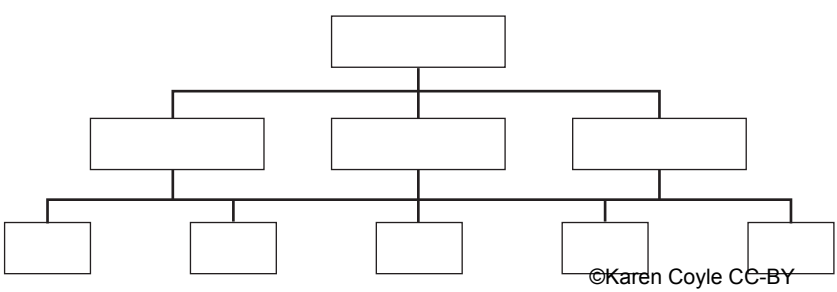
In addition, the inclusion of many-to-many relationships between expression and manifestation means that the structure of Group 1 in data modeling is by definition a network structure. However, FRBR treats the structure as "a continuous chain," as described in section 5.2.1:

> It should be noted that although the relationships between *work, expression, manifestation,* and *item* are depicted in the entity-relationship diagram in a segmented way, they operate logically as a continuous chain. That is to say that the relationship from *work* to *expression* carries through to the relationship from *expression* to *manifestation*, and those two relationships subsequently carry through to the relationship from *manifestation* to *item.* Thus when a relationship is made between an *expression* and a *manifestation* that embodies the *expression,* the *manifestation* is at the same time logically linked to the *work* that is realized through the *expression,* given that the *expression* has been linked to the *work* it realizes. (FRBR Final Report, 58–59)

It's easy to understand why people assume that there is inheritance. The statement that relationships "carry through" does sound like an inheritance model. Ignoring the many-to-many relationships (which are illustrated in the FRBR diagrams only through the use of multiple arrow-heads, and are easy to overlook in that form), you can produce an instance of FRBR that looks like that shown in figure 9.1.

FIGURE 9.1

A hierarchical model

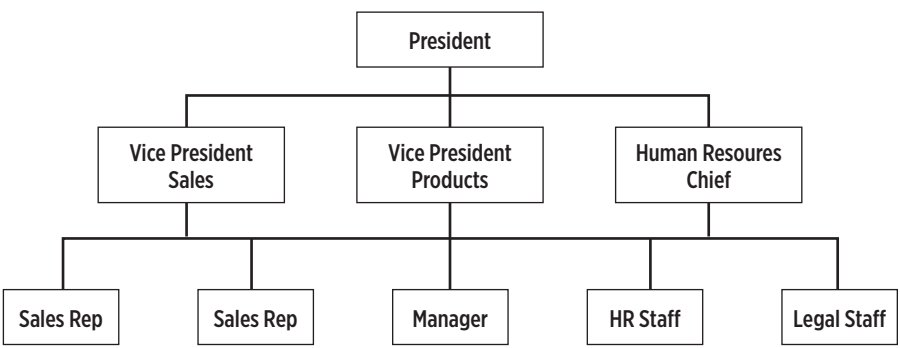


This is visually the same as a taxonomy. In a taxonomy, the lower levels do inherit from the upper levels, because each lower level is a "type of" the defined upper level. For example, a dog is a type of mammal, so dog inherits qualities from the super-class mammal. However, an expression is not a type of work, and a manifestation is not a type of expression. The structure of FRBR does not represent a taxonomy and the entities are not in a class/sub-class relationship to each other.

The shape of the diagram is deceptive. The same diagram can, for example, represent an organization chart, as in figure 9.2. An organization chart does not have inheritance—lower-level employees are not "types of" the levels above them, and do not inherit the tasks or salary of their superiors. Each employee is a separate entity with a defined relationship to the employees "above" and "below." Like the FRBR entities, these relationships are not of the "type of x" or "is an x" nature; an office worker is not a type of manager even though linked to the manager in the organization chart.

FIGURE 9.2

An organization chart



The four levels of Group 1 are not hierarchical in the common sense of that word. They do have a defined order, and the order goes from concrete (item) to increasingly abstract (manifestation, expression, work). But the entities as diagramed are independent of each other, in the same way that a person entity is independent of a work entity, with which it can have certain relationships.

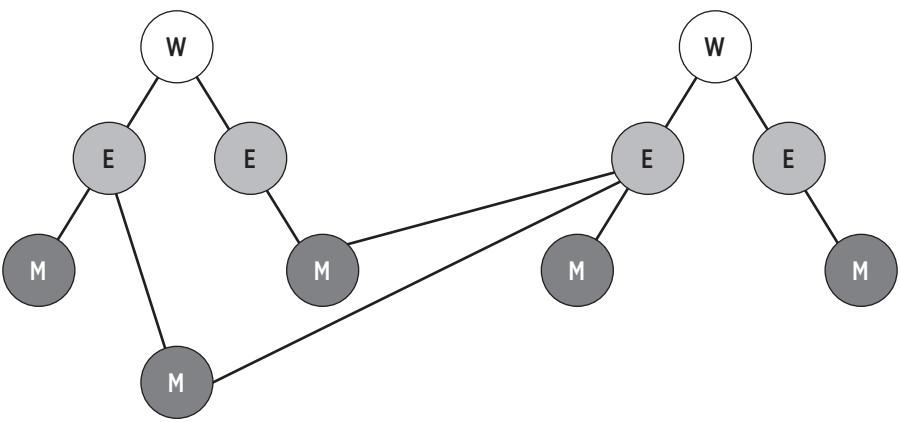
In the formal modeling of FRBR as a Semantic Web vocabulary, the Group 1 entities are defined as "disjoint." What disjoint means in that technology is that the two entities cannot share any attributes or relationships. It is for this

reason that each FRBR entity must have its own distinct set of attributes, like "work title" for the work entity and "title proper" for the manifestation entity, even though in some sense "title is a title is a title." This disjointness also means, though, that the entities cannot inherit attributes from each other because an expression or manifestation is not allowed to be described with the attributes of the work, and vice versa.

More importantly, Group 1 is not a tree structure like a taxonomy or an organization chart. There are many-to-many relationships between the Group 1 entities, which means that Group 1 is a network of relationships. A more accurate diagram would look like figure 9.3.

FIGURE 9.3

FRBR Group 1 with many-to-many illustrated



Note that the diagram in figure 9.3 only includes the primary relationships of Group 1. You might say that they are the glue that holds Group 1 together; they are structural in nature. The bibliographic relationships, like "adaptation of," "translation of," or "supplements," are also important, and presumably are part of the fulfillment of user needs in the library catalog. Adding only a few of these relationships provides a much more complex and graph-like image.

In data design, a network structure is quite common, and the many-to-many relationship is hardly unusual. It does, however, mean that the implementation of the conceptual model as a logical model must do some extra work to make the many-to-many relationships function as desired.

GROUP 1
ENTITIES AS "DISJOINT"

As mentioned above, the Group 1 entities defined in FRBR do not share any attributes. The attributes are each exclusively related to a single entity, and in the case of attributes that could be used for more than one FRBR entity, the attribute has been defined to describe only one entity, as in work identifier, expression identifier, manifestation identifier, or item identifier. Thus, each entity is a walled garden, distinct from the other Group 1 entities.

Yet this separation of entities is contradicted in the FRBR Final Report's textual description of the entities. In the text of the document, an expression is a work as expressed, and therefore includes the content and meaning of the work. A manifestation is a real world realization of the expressed work, and therefore includes the content and meaning of the work as well as the expression of the work. This alone disproves the theory that the entities are entirely separate from each other. However, as we've seen, the E-R diagram and the text of the FRBR Final Report are not perfectly aligned. The E-R diagram shows a bibliographic description that has been broken into four dependent parts that must be reassembled to create a whole. One way to interpret this is that the text describes the concepts, but the E-R diagrams represent a further analysis that pertains to FRBR as a data model. However, this does not resolve the differences between those two views.

One of the first arguments against disjointness of the FRBR entities as a choice for bibliographic data is that, as the FRBR Final Report itself allows, not all bibliographic communities are expected to have the same precise definitions of the FRBR entities:

> Because the notion of a *work* is abstract, it is difficult to define precise boundaries for the entity. The concept of what constitutes a *work* and where the line of demarcation lies between one *work* and another may in fact be viewed differently from one culture to another. Consequently the bibliographic conventions established by various cultures or national groups may differ in terms of the criteria they use for determining the boundaries between one *work* and another. (FRBR Final Report, 16)

Clearly, if the boundaries can be different, then the assignment of attributes to entities must be able to differ. However, nothing in the description of the attributes of each entity in FRBR takes this into account. Somehow, moving

from page 16 of the FRBR Final Report, which describes the work, to page 32, where the attributes are defined, all notion of possible variation is abandoned. As Robert Maxwell says in *FRBR: A Guide for the Perplexed: "*Given that FRBR emphasizes the fluidity of the concept behind the entity work, it is somewhat surprising that the document immediately gets down to the business of defining exactly where that line or boundary is (FRBR 3.2.1, 16–17)" (Maxwell 2008). The FRBR Final Report's description of the work and its list of attributes as a single, unvarying view means that any deviation from that view is a deviation from the FRBR model itself. The declaration of disjointness between entities enforces the view that there is only one "right way" to model the bibliographic universe.

One must be cautious in defining entities as disjoint because disjointness results in considerable rigidity. That rigidity makes it difficult to change ("women cannot be priests" or "two men cannot marry"), and it also makes it difficult for data from different communities that may not have the same restrictions to interoperate. Disjointness means that all users of FRBR must have the exact same definitions of the entities of Group 1 (which is most likely the purpose of the strict definition). If a specialist community determines that a physical attribute like color defines a new expression instead of being an attribute of manifestation, as defined in FRBR, they will be violating the FRBR model. This rigidity, as well as the fact that FRBR is considered "true" in its current form, means that any application must be either FRBR or not-FRBR, thus splitting the bibliographic world into noncompatible factions.

AGGREGATES

FRBR depicts Group 1 as a network in which each entity is in a specific relationship with one other of the Group 1 entities. They therefore form a continuous line from work to item, and item back to work. Difficulties arise, however, in the relationship between expression and manifestation. What often interferes here is the complication that publishers and producers of creative works add to the picture. Although it may be quite accurate to say that an expression is manifested in a physical product, it is something else to say that the physical product is solely the manifestation of the expression. The reason is that the physical, publisher-produced package nearly always has content and qualities that are in addition to the expression. From the design of the package to liner notes, creator biographies and prefatory material, the expression is packaged as a manifestation with content provided by the publisher or producer.

This additional content is described in detail in Gerard Genette's *Paratexts: Thresholds of Interpretation* (1997). Paratexts are all of the contents of a publication that are not the primary text. Some are provided by the creator of the primary content, but others are added by the product creator, the publishing house. Paratexts are sometimes recorded in library catalog data, depending on their perceived importance (e.g., a preface by a well-known scholar), but not always. In cataloging, one of the more important paratexts supplied by the publisher is the title page, which is considered the primary source of information for the catalog record. Yet the publisher is recorded only as an element of that page, and is not credited with any specific creative role for included content except in some rare books or archival publications.

Not only do publishers include paratexts in publications, they also create aggregations of expressions of works, and not always with the participation of the original creator. Aggregates can contain works of a single author or those of multiple authors. Because of this, the difference between multiple expressions in a single manifestation and the presence of an expression along with paratexts is not nearly as clear as it could be. A version of Dante's *Inferno,* with a translation into English, a lengthy introduction, and copious explanatory footnotes is hardly a mere manifestation of the expression of that work, yet there is no physical separation between the works of Dante and his commenter that could easily lead to the definition of two separate works. Some collections of seemingly separate works may in fact have the intellectual characteristics of a single work, as in publications of a professor's lecture series. On the other hand, a recording of two or three musical pieces by a single composer may be little more than the publisher's determination of an optimum number of minutes given the physical medium of publication. A publication of separately authored papers may in some cases cover a wide range of topics, making "workness" nearly impossible to define for the whole. But in other cases the separate essays may have a coherence that is hardly different from a group of chapters written by a single author.

The question that one asks at this point is whether the FRBR manifestation refers to the expression that has been published, or whether it refers to the entire published package. The FRBR Final Report allows for either the treatment of an aggregate as a work in itself, or as a whole/part relationship, such as with individual chapters in a book or articles in a journal. The whole/part relationship has two types: the dependent part, which does not stand alone or is integral with the whole (such as illustrations that accompany and support a text) or the independent part, such as an article in a journal. The FRBR Final Report concludes

that although "there often will be no reason for a dependent part of a work to be separately identified or described . . . independent parts of a work are more apt to be identified and described in their own right." (FRBR Final Report, 68)

If a FRBR manifestation does not refer to the entire publication, then it isn't clear how the fullness of the publication is covered in that model. It also isn't clear what the difference is between "is manifested as" and "is part of." FRBR includes part/whole relationships in its model. This, however, is where things get complex.

Because the bibliographic description is separated into the four Group 1 entities, WEMI, it is implicit that all creative output has this structure. Thus, an introductory essay in a published book also should have the four-part nature of work, expression, manifestation, and item. One could take that further and consider each paratext and each design element of the finished publication to have "WEMI-ness." A book cover design surely would be considered a manifested expression of a work to the extent that any other visual resource is, and the same could be said for photographs or illustrations within the book.

In this way, the layered bibliographic description that is WEMI complicates the part/whole relationship, because one has to decide where along the cascade of abstractions of WEMI the part/whole relationship exists. Is this part of a work? Part of an expression? Part of a manifestation? These are not easy questions, as the FRBR Study Group discovered.

The FRBR Study Group created a subgroup to study the complications brought on by aggregate publications. The Working Group on Aggregates narrowed down the potential solutions to aggregations to two competing views: aggregates as manifestations, which was authored by Ed O'Neil and Maja Žumer, and aggregates as works, by Barbara Tillett.

The argument by O'Neill and Žumer (undated) on aggregates as manifestations makes use of the allowed many-to-many relationship between expressions and manifestations in FRBR. Because any manifestation can manifest multiple expressions, an aggregate is a "combination of expressions" in a single manifestation. They also note that there are different types of aggregates: collections (which are aggregates of independent works), augmentations (which contain supplemental material around a primary work or works, like the aforementioned translation and commentary on Dante), and parallel aggregates, which are often texts that appear in more than one language in the same publication. An aggregate as manifestation has separate expressions of works that are manifested in a single manifestation.

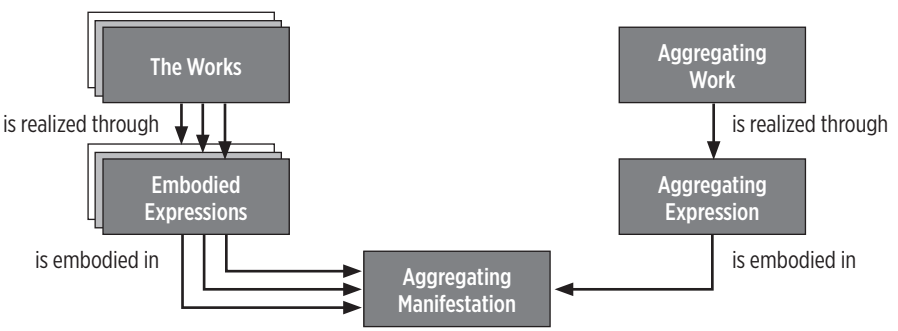
Tillett's paper (undated) takes the view that an aggregate is a work in itself, and that it has a whole/part relationship with any works that are included in the

aggregated work. Therefore, a "work can be a work of works," which could be shown with a recursion symbol on the box for the work entity. She states that "The recursive symbol was omitted from the final text, but because this is causing confusion, I feel it should be reintroduced."

The final report of the Working Group on Aggregates (O'Neill et al. 2011) concludes with an aggregation of its own, a combination of features from the two viewpoints described above. The group concluded that the FRBR approach to aggregates would use the many-to-many relationship between expressions and manifestations. However, because aggregates have some properties of their own (such as a creator of the aggregate), aggregates may also be seen as separate works. Any one instance can have multiple works as well as a single aggregating work, as shown in figure 9.4.

FIGURE 9.4

The solution to FRBR aggregates



The recursive element suggested by Tillett was not included in the FRBR Final Report. The Final Report also did not resolve the difference between a manifestation that manifests multiple expressions and a manifestation with parts. This seems to be a significant weakness in FRBR that needs resolution but may require more experience with the model.

Heidrun Wiesenmüller (2011) of Stuttgart Media University studied the FRBR aggregate model and questioned the logic of the sub-group's conclusion. In her view, the treatment of both individual works and aggregating works as FRBR works at the same level did not accurately reflect the nature of those works. Her suggestion was that there should be two levels of work, one for the aggregate and one for the individual works. This, however, produces an even more complex

set of relationships, at least in terms of how they are diagrammed. However, no model that includes aggregations will be simple.

One solution not considered by the Study Group on Aggregates would be to consider that all manifestations are, essentially, aggregates, comprising at least a primary expression of a work and the paratexts accompany that content. Taking this view would lead to a consistent treat of all manifestations since a publication always goes beyond the mere manifestation of an expression. Even the choice of page design, the use of covers, employing a font and pagination, perhaps adding a table of contents and index—all of these make the manifestation more than a manifestation of an expression. The alignment of FRBR with the CIDOC resource model, FRBRoo, emphasizes the publication as an event in the chain from creator to audience, and therefore injects the act of publication as external to the creative workflow. However, FRBRoo very definitely defines aggregates as works and does not have an equivalent to FRBR's aggregation as manifestation. In the view of CIDOC's resource model, the FRBR manifestation is a compound of content and carrier that needs to be broken apart. The work on FRBRoo may be the best analysis of why the FRBR Study Group was unable to find a solid solution to the problem of aggregate publications.

Before concluding, it is important to acknowledge that decisions made for the purposes of library cataloging are not intended to solve philosophical questions about the nature of reality. The catalog has practical goals relating to inventorying the library collection and serving those who seek materials in the library. For that reason it is legitimate for library cataloging rules to make decisions that serve the library's need, even if one could argue that they are not somehow philosophically sound. When the members of the Working Group on Aggregates, a subgroup of the FRBR Study Group, were tasked with formulating a solution to the problem of how to deal with the part/whole and aggregates issue, their focus was rightly on solving this problem for library cataloging. Whether they did so in the best way possible is something I do not feel qualified to judge. It is unfortunate, however, that this problem did not lead to a reevaluation of any of the assumptions on the nature of the FRBR entities, in particular on the relationships between the expression and the manifestation. One possible adjustment could have been to redefine the manifestation separate from the package or container. The manifestation of the expression would then be something contained within the publication. This would have separated the publisher's package from the creative work it encompasses and made it possible to describe the two independently. Instead, the manifestation arises solely from the expression and has no individuality of its

own, leaving all of the effort of publishers to be treated as a kind of afterthought. I doubt if publishers would find that characterization flattering.

The study of aggregates is evidence of a significant difference between the cataloging view of library metadata and the view that a data designer might take. It is quite awkward (and not recommended) for data designs to have exceptions or "either/or" situations. The study of aggregates as manifestations cites statistics from OCLC showing that aggregates are only a small portion of the bibliographic universe. However, in data modeling, it does not matter if 2 percent or 98 percent of your instances will exhibit the characteristic in question; the model must solve the problem in a way that is valid for all of your data. For example, in a situation where you can have many-to-many relationships in your data, the data must be modeled that way even though a majority of your instances might be one-to-one. If there is the possibility to further develop the concepts first presented in FRBR, the treatment of aggregates is an obvious area that will need additional study.

TEN

BIBLIOGRAPHIC DESCRIPTION AND THE SEMANTIC WEB

Computer technology evolves rapidly, with change happening faster than most humans can follow. The technology of data management evolves less quickly than some other computer technologies because it often requires a rather costly conversion of data and interfaces to take advantage of new capabilities. For this reason, few can afford to jump on a new data technology bandwagon when it first appears. Instead, most wait until such a technology has matured before adoption.

The technology of the Semantic Web, although new to many, has been around for almost as long as the FRBR Final Report. The first version of RDF, the basis for the Semantic Web, was published by the World Wide Web Consortium in 1999. Adoption of RDF has been slow, but as of 2015 there is a strong movement toward implementation of this technology and integration of data with the basic functioning of the web. Most relevant to those of us in libraries, it is being used or is in the planning stages for a significant number of library and archives applications.

There is a logical progression from entity-relation modeling, such as is used in FRBR, and Semantic Web technology that models data as things and relationships. That there is a logical progression does not mean that these technologies are the same. In fact, as described in chapter 3 of this book, there are significant differences between the meaning behind entity-relation models and the Semantic Web approach that absolutely must be understood when making the transition from one to the other. But in general, an entity-relation model is not a bad predecessor to a Semantic Web data design, as long as suitable adjustments are made.

ENTITY-RELATION MODELING AND SEMANTIC WEB MODELS

> This approach differs from some other approaches in that it begins with an abstract of conceptual schema of the domain or universe in question. The universe is characterized in terms of the entities in it and the relationships that hold among them. As such, the conceptual schema is not restricted by the capabilities of any particular database system and is independent of any particular record definition. (Tillett 1994)

To understand the Semantic Web model, we need to review how data was managed before the Semantic Web, and how this relates to efforts to define bibliographic data for the Semantic Web that uses FRBR as its conceptual basis. As we've seen, FRBR was designed around an entity-relation analysis. The entity-relation method was developed to support relational database management technology, which was revolutionizing data management in the 1980s and 1990s. The problems addressed by entity-relation design are primarily related to the efficiency of storage and retrieval of data held in database management systems. An entity-relation analysis normalizes data gathered for business functions with a goal of storing each data point only once, and allowing the combination of atomized data elements to support a wide variety of business functions. At the conceptual level, entity-relation analysis identifies the primary entities that the enterprise must manage as data. Subsequent steps define data elements supporting or describing each entity, and which are needed to enable the tasks and workflow of the enterprise. These analyses may support static or dynamic processes, depending on the needs of the organization.

The Semantic Web has an entirely different approach from that used by relational databases. These differences arise from primary assumptions of the closed world of databases versus the open world of the web.

Databases reside generally within a system that is contained within the local network of an organization. Although that local network most likely connects in some way to the Internet, the database itself is not on the web. In fact, in some cases, like banks, hospitals, and government agencies, the security of the database is one of the key requirements of the data management function. For example, the banking industry has its own secure network for the exchange of information between banks; this information is not available over the open web and cannot be directly accessed by other than local users with particular privileges.

Even for those for whom the privacy and security of the data are not paramount, databases do not communicate directly with the web; the interaction with the web is managed through applications that have controlled and secure access to the database contents. This is true for library systems and their online catalogs. Although one can access the online catalog via a web browser, the data resides in a database on a private server.

In a Relational Database Management System (RDBMS), the data elements allowed are controlled by the database structure definition. This structure is made up of tables that consist of rows of data. Each data element in a table is defined as a specific data type, such as text, integer, or structured data. The data stored in the database is not in the same form as it appears in input or output actions. Instead, data in the tables is combined to present views of the data for input and output. For example, one common type of data in a database is that of names and addresses of customers, users, or employees. The data that makes up a complete mailing address may be stored across several tables such that city and state names are not repeated in the database but can be linked to the particular address when requested. The database design does not represent a record, but is a store of data elements that are input and output in different combinations, each of which may be considered a record by a particular user. A database may serve a wide variety of organizational functions, including some that do not share any data elements among them. In a large organization, management of the database is under the control of a technical department, although input and output may take place throughout the organization by different types of employees. These employees may be entirely unaware of the nature of the whole database, and only see the portion that is relevant to them through a user interface.

A common aspect of relational databases is that they have features built in for data quality control. In fact, the database design itself forces data to conform to rules. For example, certain data elements can be required in the database design. If the database has only one storage point for a data element, more than one element simply cannot be added. In particular, no new element types can be added by database users until changes are made to the database design. Because

these changes often require a significant amount of planning and testing, data stored in databases has a tendency to change only with a certain effort. Anything not included in the RDBMS design is not available to the applications that will use the data.

It is common to hear the web of linked data referred to as a way to use the web as your database. Although this analogy holds, the difference between the web as database and the closed, controlled world of the enterprise database management system is enormous. The open world of the Semantic Web has no predefined boundaries. The maxim used on the Semantic Web is that "anyone can say anything about anything." The Semantic Web has no concept of quality control analogous to an RDBMS. This has advantages as well as some challenges. The advantage is that new data types can be added as needed. Rather than the slow and often painful process of database modification that traditional closed databases require, the web of linked data can allow new concepts to be added by anyone at any time. Any data on the web can be linked to your data, just as today anyone on the web can add a document that links to a web page of yours. There is also no way to require that certain data always be present in the Semantic Web model. Because the open world of the web is considered to be ever-changing (as is the real world of information), it is not possible to predict what data will be available at any given moment in time. Data that is missing at one moment may be present a moment later. Therefore the Open World Assumption, one of the fundamental tenets of the Semantic Web, does not treat missing data as an error but merely as something momentarily unknown.

In the pre-Semantic Web data environment, we generally consider ourselves to have data encapsulated in records outside of the database. The database itself has tables and does not conform to the form of the records that are input and output. However, for those who create and use the data, the record is the unit that defines an appropriate grouping of data elements. Like a database, the Semantic Web does not have records. But the input and output, when using the Semantic Web standards based on RDF, also are not packaged as records. Regardless of the view that one has at any given time, the Semantic Web is always composed of three-part statements called triples, which are autonomous atomic statements that can be combined in any desired configuration. A search against the Semantic Web returns some number of triples that match the search query.

FRBR, and subsequently RDA, were modeled as one would model data for an RDBMS. As part of the development of RDA, the Joint Steering Committee for RDA produced a document that presented three high-level database models for bibliographic data (Delsey 2009). The most advanced of these models, and

presumably the preferred one, was called "Scenario 1," and made use of concepts from relational database design. The documents reflect some assumptions about systems efficiency and user service capabilities that were held by members of the committee:

> The data structures used to store the data and to reflect relationships, however, will have a bearing both on the efficiency of data creation and maintenance, and on the ease and effectiveness with which users are able to access the data and navigate the database. (Delsey 2009)

There are two ways in which efficiency of a database is commonly measured: the time needed for operations to the database to complete, and the space needs to data storage. In practice these two considerations are often in conflict. It is assumed that the storage requirements for FRBR work data (and to a lesser extent FRBR expression, because it contains fewer fields) would be reduced because there would be less duplication of these elements in the database. None of the statistics that we have on hand, however, measure data storage, so it remains an open question whether there are significant savings.

In terms of time necessary to complete operations, the efficiency of a database often depends on how many joins and reads are required. Joins and reads are what bring together related data that is stored in the separate database tables. Some efficiency that is gained by reducing redundancy is lost when multiple tables must be included in a single database activity. Database models are adjusted to meet these efficiency requirements by testing. This type of testing would be needed before any definite statements could be made about a FRBR-based database model.

Suffice it to say that at this time there are no studies available that support any claims of efficiency for RDA as a FRBR-based bibliographic model. How efficient the sharing of works and expressions for catalog maintenance depends on the extent of redundancy of entities in the local catalog, which is usually a function of a library's mission and size, as well as the extent to which bibliographic data is held in a shared data pool rather than copied to the local library system.

The Semantic Web model is also seen as holding promise in terms of data sharing. The Semantic Web is by its nature a shared "cloud," although the same technologies are used in closed enterprise systems. The primary data structure of the Semantic Web, RDF, is already being used in large corporations that need to share and link data between operations and over large geographical spaces. There is tangible tension between the Semantic Web design for the open web

and the use of these technologies in closely held data environments. For many enterprises, it's not a matter of entirely open or entirely closed but, as with traditional databases, using the networking capabilities of the Internet to share select information with a wider public.

FRBR IN RDF

The conceptual model of FRBR and variants on that model have been expressed in recent RDF vocabularies developed for library data. FRBR has also been used in bibliographic models designed outside of the library. There are also Semantic Web implementations of bibliographic data, such as some develop for academic citations, that do not make use of the FRBR conceptual entities. This section focuses on some of the projects that have transformed the entity-relation model of FRBR to a Semantic Web vocabulary.

FRBRer

In 2011, the IFLA FRBR Review Group, the group that now maintains the FRBR standards (including functional requirements for authority and subject data) issued its official version of FRBR in RDF. Called FRBRer, with the "er" standing for entity-relation, this version is based on the 2009 edition of the FRBR Study Group's final report.

FRBRer is an encoding of FRBR as defined in the document. FRBRer uses the World Wide Web Consortium's standard for the definition of vocabularies, the Web Ontology Language (OWL), which is described in chapter 3 of this volume. As the literal translation of an entity-relation model, FRBRer does not follow some practices that are common in Semantic Web vocabularies. In particular, FRBRer does not make use of super- or sub-classes to define logical types of entities. Without class relationships, which are heavily used in Semantic Web vocabularies, FRBRer cannot define a single relationship that is valid between a work and any members of FRBR Group 2. Instead, FRBRer must develop a specific relationship for each individual Group 2 entity: for example, "is created by person," or "is created by corporate body." This must be done for each of the relationships involving Group 2 entities such as "realized by," "produced by," and "owned by." The later addition of "family" to Group 2 meant that new relationships had to be added for it as well. This has an effect on the ease of extensibility of FRBRer: with class relationships it would only be necessary to define the entity "family" as a sub-class of the class "responsible entity," and family

would be immediately usable with all relationships defined between Group 2 and Group 1. As it is, the addition of family also requires the family-specific definition of each of the relevant relationships to the other entities.

It is clear that FRBRer is designed with a closed-world point of view rather than the open view of the Semantic Web. This view in inherited from the entity-relation origins of FRBR, because entity-relation modeling is a design method for database technologies and not for the Semantic Web. As I noted in the chapter on technology, there is a common misinterpretation of the Semantic Web vocabulary definition language OWL that reads the OWL rules as quality-control constraints on the data rather than as axioms for making inferences from the data. FRBRer is based on this misinterpretation.

One way that FRBRer affirms its closed-world view is by declaring all entities and all of the FRBR attributes as disjoint from each other. This means that it isn't possible for anyone using FRBRer to create data that varies in the assignment of attributes to entities. The result of this is that the data coded in this vocabulary is very fragile, with any deviation from the defined terms causing OWL-aware software to fail to function as desired. This fragility is not easy to mitigate because OWL does not have any functions that enforce quality control on data; instead, in the open world where "anyone can say anything about anything," vocabularies need to be as forgiving as possible. The strictness implied by disjoint classes may be unrealistic in an imperfect world, with a few exceptions for undeniable truths, such as "up" being disjoint "down." Nearly any category has exceptions that need to be handled, even the fundamental states "solid, liquid, or gas," which include intermediate transitions that can be more than one state at the same time. In the open world of the web, the strictness implied by disjoint classes generally prevents this data from interoperating with any data that is based on a different model, even if they have linkable elements in common. In the case of FRBRer, this creates an incompatibility with any bibliographic data that does not precisely separate its bibliographic description into identically defined entities. Not only does this mean that variants of FRBR as desired by some library specialist and non-book communities cannot be used alongside conformant FRBRer, but that FRBRer on the open web may not be able to link to similar data using models like FRBRcore, BIBFRAME, or even RDA in RDF.

In this way the primary advantage of the Semantic Web, discovery across heterogeneous data contributed to the web of data by different communities, is negated by the definition of vocabularies using OWL with an inappropriate closed-world assumption. FRBRer, as defined, is a vocabulary that will necessarily be usable only in its own silo.

RDA in RDF

Resource Description and Access (RDA), the cataloging rules adopted in 2013 by many North American libraries, also has a defined RDF-based vocabulary. The development of RDA was closely tied to the bibliographic concepts presented in the FRBR Final Report, and the structure of the RDA rules follows that of the FRBR groups—primarily Group 1, which is the main focus of descriptive cataloging. A list of data elements, each with their assigned FRBR entity, was published as a supplement to the final text of RDA. After a 2007 meeting between volunteers active in the Dublin Core Metadata Community and members of the RDA development group, the RDA elements were defined in RDF on an experimental basis. This activity preceded the development of FRBRer, and RDA in RDF was based on a very loose vocabulary of FRBR entities as RDF classes. Although FRBRer is now available, and there is an official version of RDA in RDF that is managed by the Joint Steering Committee (JSC) that maintains the cataloging rules, the JSC has so far chosen not to follow the vocabulary as defined by FRBRer. Instead, RDA in RDF has created its own FRBR classes to define the domain of each of the data elements, but does not attempt to impose E-R-like rules on the resulting bibliographic description. Each RDA element is also sub-classed to a parent element that is not associated with any FRBR entity, creating a FRBR-neutral version of the vocabulary that may be more acceptable to nonlibrary communities for whom the Group 1 entities of FRBR are unfamiliar, and perhaps even not useful. RDA in RDF creates a super-class for person, corporate body, and family that is called "agent," but no super-class for the Group 1 entities. Recall that the FRBR Review Group has made clear that in their analysis the groups are not to be represented as super-classes; instead, they exist as organizing elements in the documentation only.

With an awareness of the open world and the fact that there are many sources of bibliographic data, both within the library community and outside of it, RDA in RDF proposes linking RDA elements to commonly used terms from other vocabularies using sub-class relationships, with a more general vocabulary like Dublin Core Metadata Terms as the common language. This can bridge the gap between the over 900 elements in RDA and bibliographic data as approached by nonlibrarians. There probably is no other community that has so many different types of titles (key title, parallel title, series title, parallel series title, etc.), yet most communities creating bibliographic data will have an element that is compatible with Dublin Core "title," or that can be sub-classed to it. Because the primary goal of the Semantic Web is to allow linking between comparable data across the web, it is a good idea for everyone to design links into their community-specific

vocabulary to well-known vocabularies used on the web. In this way one avoids being stuck in a silo where library data can only connect to other library data on the web.

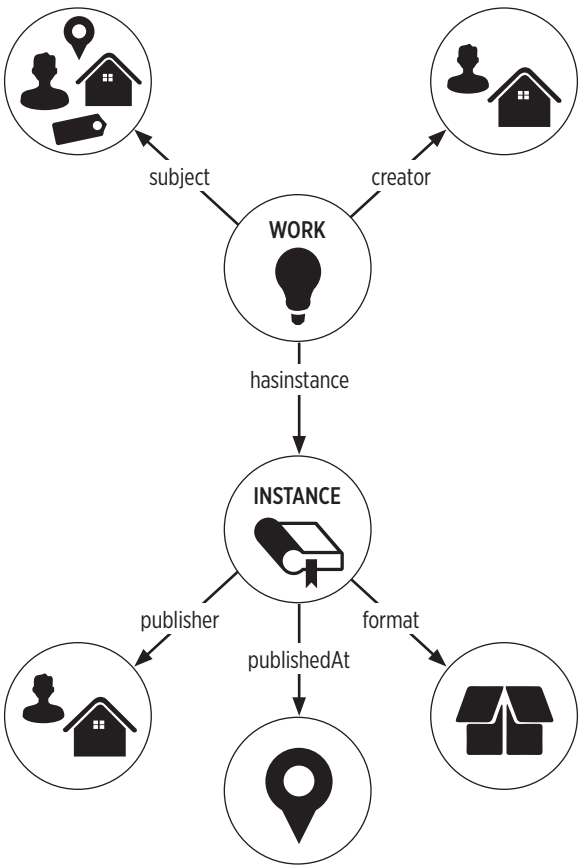
That said, RDA in RDF itself has few relationships between terms, and few uses of classes and sub-classes. This is possibly a reflection of its origins as a list of terms derived from the thirty-odd chapters of the RDA cataloging rules. You could say that RDA has not yet been subjected to the data design phase that would look at the desired functionality that could be delivered with complex bibliographic data expressed in RDF. This type of design often begins with a statement of use cases: what is it that we want to do, and that we could do, with this data? Recall that in chapte 2 on modeling we had simple use cases like "find a book of which the author is known." Given the kinds of capabilities that we have with current data management technology, including the possibility of using the entire web as the context for our information services, the number and kinds of use cases would surely grow. Do we want to promote linking from online sites for readers, like GoodReads or LibraryThing, to library materials? From the music database MusicBrainz to library collections of recorded sound? From the Internet Movie Database to film descriptions in library catalogs? Do we want to incorporate more information about authors alongside the library's bibliographic holdings? Do we want to continue to store most catalog data locally, or would "cloud-based" sharing of data be more efficient? The list would end up being quite long, but without an exploration of goals, we cannot make rational decisions to guide our development.

BIBFRAME

The RDF vocabulary of BIBFRAME, which is being developed simultaneously by Library of Congress, Zepheira, and a number of library projects, is not an implementation of FRBR, but is clearly influenced by the FRBR model. It has a two-entity model of bibliographic description, with the entities called work and instance. The BIBFRAME work represents the content portion of the bibliographic description, and the instance describes the carrier. Generally speaking, the BIBFRAME work encompasses attributes that are associated with both the FRBR work and expression; the BIBFRAME instance is analogous to the FRBR manifestation. Item-level information is not treated as one of the primary bibliographic entities in BIBFRAME. BIBFRAME also resembles the pre-FRBR data model with a central bibliographic description plus authority entities that are similar to name and subject authorities (figure 10.1).

FIGURE 10.1

BIBFRAME model



BIBFRAME and FRBR come out of different communities within the library environment. Although the FRBR Final Report provides background on the bibliographic theory that led to the creation of FRBR, the link between standard cataloging practice and BIBFRAME is less clear. As the "named successor" to the current standard bibliographic record, MARC 21, BIBFRAME is primarily a data standard. To accommodate MARC 21, BIBFRAME will necessarily have more detail than the FRBR conceptual model, which was developed ostensibly as a minimum set of bibliographic data.

SEPARATE WORK AND INSTANCE WITH LINKS	SINGLE RESOURCE WITH WORK AND INSTANCE DATA ELEMENTS
ex:ResourceA bf:workTitle ex:AdventuresOfTomSawyer ; bf:hasInstance ex:ResourceB ; bf:creator lcna: n79021164 ; bf:language is0639–2:eng . ex:ResourceB bf:providerDate “1996” ; bf:instanceOf ex:ResourceA ; bf:instanceTitle ex:TheAdventuresOfTomSawyer .	ex:ResourceA bf:creator lcna: n79021164 ; bf:workTitle ex:AdventuresOfTomSawyer> ; bf:language is0639–2:eng ; bf:instanceTitle ex:TheAdventuresOfTomSawyer> ; bf:providerDate “1996” .

BIBFRAME’s vocabulary is more compatible with the open web and with potential variations in bibliographic concepts than is FRBRer. BIBFRAME makes use of classes and sub-classes in ways that are convenient for designers of systems, which will probably facilitate searching and other system functionality. It does not define any classes as disjoint. Because of this, both of the above examples are valid instances of the BIBFRAME vocabulary, with no loss of information:

The above examples illustrate that the storage of descriptive elements in separate work and instance units is not required in BIBFRAME. Whether you keep work and instance separate or not can depend on your needs. In fact, at times data can be stored without separate works and instances, yet can be transmitted as separate BF entities when desired because the entities can always be created from a properly defined RDF vocabulary. This is an important lesson about RDF, especially as compared to the data models that we are most familiar with: in RDF, the meaning of an element is in the defined vocabulary, not in a record structure. You can define data that adheres to the concepts of work and instance, or even of work, expression, manifestation, and item, without that predetermining the structure of your data. Whatever view of your data you work with at any given moment depends only on what works best for you for that function.

ALTERNATIVE MODELS USING FRBR CONCEPTS

In previous chapters we looked at how the reality of FRBR is different from its stated goals: both the goals that led to the creation of the FRBR Study Group and the goal presented in the FRBR Final Report itself. Yet there is no question that FRBR, and in particular the Group 1 entities, resonates with many people

both within librarianship and outside of it. We see a general recognition that the bibliographic resource is a complex thing that can be approached from a number of different points of view. This, at least, seems to garner wide agreement. What is not agreed upon, however, is a single interpretation of that complexity. It seems that many have a need to express the complexity of what they describe bibliographically, but those needs have a great number of potential expressions.

Those who have made use, in their own ways, of FRBR concepts have employed them to reinterpret the bibliographic entity as a cascade of abstractions, from the most abstract work through some number of intermediary levels until the actual physical item is described. The number of variants of this path, however, seem to be without limit, nor do the multi-entity models even agree on the nature of the levels of abstraction. FRBR itself is a shallow model, with a set of entities but no sub-entities. In some cases, those who borrow FRBR concepts extend the model significantly beyond its flat nature. In other cases, the model is reduced in the number of entities while at the time it is given greater depth through the creation of a hierarchy of entities.

Because it is a conceptual model, there is nothing in FRBR that has not been treated as open to reinterpretation, not even the seemingly uncontroversial concept of physical item. Where one draws the line between the physical and bibliographical is not as clear as you might think. FRBR, the library, and archive-based models often combine physical description (e.g., dimensions of the package) with a description of what is printed on the package itself. This is a mixture of information that harks back to the catalog record that presents a single line or field that includes extent, illustrative matter, and size ("xiii, 368 p. : ill. ; 24 cm.") Arguably a more logical separation would focus solely on physical properties. However, defining what is relevantly physical regarding digital resources, is not easy, and is even more difficult in a mixture of hard copy and digital resources,.

Each of the models presented below is a response to FRBR with some alterations. Given that FRBR presents itself as a conceptual model (at least in the text of the document), each of these uses the conceptual model and proposes a direction or implementation of it under different assumptions or using different technologies. Included here is just a selection of models that have riffed on FRBR's melody, or have come to a similar conclusion independently. Each of these confirms some aspects of the multi-entity bibliographic model, while drawing into question some others.

FRBRcore

The first development of the FRBR conceptual model in RDF was done in 2005 by Ian Davis and Richard Newman, in an effort that was not supported by the

IFLA group. This version of FRBR was called "FRBRcore" because it included the entities and relationships of FRBR but not the attributes. This may seem odd, as the attributes seem to have the greatest importance because they carry the descriptive information about bibliographic resources. This approach, however, makes sense if the role of classes in RDF is understood. RDF classes are similar to entities in a conception entity-relation model, and therefore entities are often designed as classes when models are interpreted in RDF. The classes in FRBRcore provide a conceptual framework that can be extended as needed. As we'll see when we discuss alternate models based on FRBR, a single conceptual framework can be used as the basis of some very different solutions. FRBRcore makes those solutions possible.

Davis and Newman were familiar with library models but they also were involved in RDF implementation. Although they were the first to develop an RDF version of FRBR, their vocabulary was never endorsed by the IFLA Working Group on the Functional Requirements for Bibliographic Records. Indeed, FRBRcore varied some from the E-R model in the FRBR Final Report. Davis and Newman added a super-class that encompasses the entities of Group 1, which they called "endeavor." They also created super-classes for Groups 2 and 3, "responsible entity" and "subject," respectively. In addition, they assigned a super-class, "spatial thing," to both object and place, because both of these could require geo-location attributes. The class structure in RDF is used in applications to address data at different levels of specificity. For example, the class "responsible entity" becomes shorthand for "person or corporate body" in program functions. The total number of entities in FRBR are relatively small, but in many RDF vocabularies the classes outnumber the descriptive elements, and are vital for efficient processing of the data.

FRBRcore could be considered experimental in nature, and is definitely not intended to be complete. The authors appear to have added a few elements intended to illustrate how FRBRcore could be extended. As sub-classes of work, they defined ClassicalWork, LegalWork, LiteraryWork, and ScholarlyWork. Each of these is a type of the more general concept of work. This extension of the work in FRBRcore may be surprising to those accustomed to library bibliographic data, because nowhere in FRBR or in any of the cataloging rules is there a discussion of types of works similar to those devised by Davis and Newman. The nearest thing is the FRBR attribute form of work, which includes as examples "play, poem, and novel." FRBR form of work isn't the same as the FRBRcore subtypes of work; however, either or both could be the basis for extending the work concept. In fact, a concept like work could be extended in a number of different directions, because in RDF classes are not exclusive. There is also no limitation in RDF on

the creation of a super-class that is a generalization of one or more classes. That FRBRcore has the super-class "endeavour" does not mean that the FRBR entities of Group 1 are no longer valid or that their meaning has changed.

FRBRcore is highly visible today in the linked data cloud. The Linked Open Vocabularies project reports that FRBRcore is used in twelve vocabularies, and appears in nearly 30 million instances, although most uses are from the union catalogs of Bavaria, Berlin, and Brandenberg, and the union catalogs of Hessen and parts of the Rhineland. These catalogs account for about 24 million instances of FRBRcore in RDF triples.

FRBRoo's Object-Oriented Model

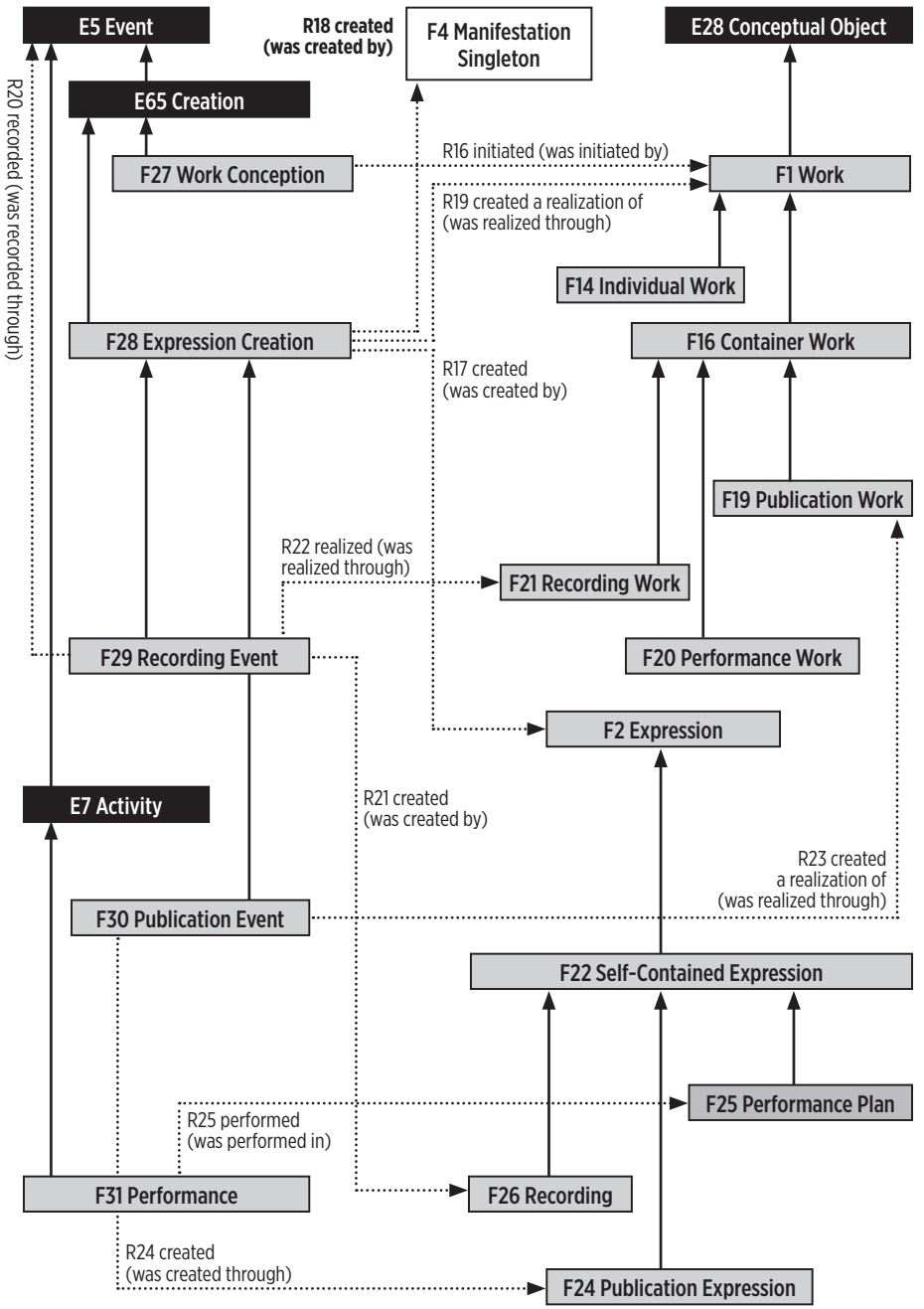
FRBRoo is a harmonization of the FRBR entity-relation model and the Conceptual Reference Model of the International Council of Museums (CIDOC CRM). It is, as its name indicates, an object-oriented data model, but the vocabulary has also been defined in RDF. FRBRoo is not intended as a replacement for FRBRer, but as an interpretation for an object-oriented environment and a harmonization with museum practice. The FRBRoo document explains that its project also serves as a proof of concept of the FRBR model:

> Expressing the FRBR model in a different formalism than the one in which it was originally developed provides a means to evaluate the model in terms of its internal consistency. It is also a good opportunity to correct some semantic inconsistencies or inaccuracies in the formulation of FRBR that may be regarded as negligible when FRBRER is only used in a library catalogue context, but that prove to be quite crucial from the moment one strives to design an overall model for the integration of cultural heritage related information. (Doerr 2006)

Object-oriented models define both things and processes, while the E-R model used in FRBR is a static definition of entities and relations. However, some of the relations in FRBR (e.g., "manifests" or "expresses") imply some action, and FRBRoo has taken an action or event-oriented view. FRBRoo includes an event for each of the WEMI concepts, such as Work Conception, Expression Creation, Carrier Production Event, and Publication Event. It recognizes the importance of these events and the role played by the various actors that are involved in the transformation from a creative concept to something shared with others.

FIGURE 10.2

A segment of the FRBRoo model—From Work to Expression, dynamic view



> FRBRer envisions bibliographic entities as static, ever-existing things that come from nowhere, and overlooks the complicated path from the initial idea for a new work in a creator's mind to the physical item in a user's hands through the dramatically important decision-making on behalf of publishers. As a matter of fact, bibliographic records do contain implicit information about that complicated path and the relationships it implies between and among bibliographic objects; FRBRoo digs that implicit information out of bibliographic structures, e.g. the precise meaning of "date of publication." (Doerr 2006)

FRBRoo also extends the FRBR entities using super- and sub-classes, as can be seen in figure 10.2. For example, there is an abstract class super to work and expression called "conceptual object," and sub-classes under work itself for types of works, such as "individual work," "complex work," and "publication work."

Significantly, FRBRoo recognizes publication as an action, and treats Publication Work as an entity sub-classed to Container Work. This at least partially responds to the difficulty that the FRBR Review Group had with aggregates, and smooths the transition from the expression to a publicly available package that has physical and intellectual characteristics that are added by a publisher but that are not included in the expression itself. In FRBRoo, every publication is a kind of aggregate, because it always contains some aspect of creation added by the publisher. This makes a publication, by definition, a package of multiple creation activities. Works that are not published, such as works of art, can be manifested without the intervention of a publisher, and therefore do directly manifest what has been expressed. In the 2006 version of the FRBRoo documentation, edited by Martin Doerr and Patrick Le Boeuf, published works were likened to car models in their relationship between the original creator and what comes out of a manufacturing process.

> "Manifestation" can be two completely different things: Either it is an industrial product, i.e., a Type, like a particular car model, or it is a Physical Man-Made Thing that was produced as a unique carrier of an Expression. Industrially printed books belong to the first category, and are indirectly related to the main author's original creations. (Doerr 2006)

This echoes some of my own concerns about FRBR's treatment of the manifestation as being in a direct line from the expression of the work, without recognition of the many aspects of the published resource (book, sound recording, or film) that are contributed by the publisher.

> In the course of discussion however it was recognized, that virtually any book is composed of multiple, distinct works: the text, the illustrations, the editors work on lay-out, type phase etc. The latter was widely ignored in FRBR. . . . This situation demanded for a general model explicating both the individual contribution and the unity of the integrated product. (Doerr 2007)

FRBRoo appears to have a consistent grounding in object-oriented technology, and to be shepherded by a group that understands that technology. That said, FRBRoo is extremely complex, and understood by few. If it is to be adopted it will need a user-facing interface that is comprehended by a wide range of metadata producers.

<indecs> Event-Oriented Model

The <indecs> metadata model was developed in the late 1990s, and the current revised version is from 2000. <indecs> models metadata for intellectual resources around an e-commerce viewpoint, but intends its design to interoperate with the full product flow from the original creator, through the publishing and manufacturing steps, to outlets like stores and libraries, and then to the end-user.

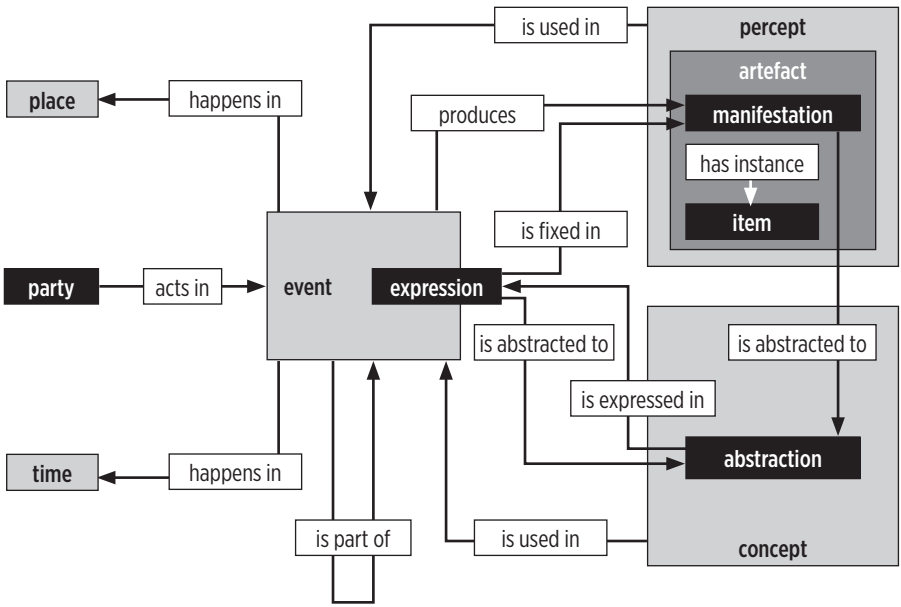
<indecs> takes an event-based view of the metadata model, and defines metadata as a "relationship that someone claims to exist between two entities." All relationships are events, either static or dynamic, and events have inputs, agents, and outputs. The agents are key in the <indecs> model because these inform the question of intellectual property rights, which are one of the main elements of commerce.

The treatment of each entity as being the result of an event with the involvement of an agent resolves some of the questions about the nature of the FRBR Group 1 entities: where FRBR has works, expressions, and manifestations as primary entities with flat relationships and no intervening human activity, <indecs> includes the persons or agents that act to create the entities (figure 10.3). This answers the question: how does a work get expressed, and by whom?

<indecs> also recognizes that different media and different products can have different events. In particular, the performance event for music produces types of expressions that are not common for textual works. <indecs> places no limits on the types of relations that can be included in the model, although it lays out a general framework of metadata properties.

FIGURE 10.3

The <indecs> model of bibliographic metadata



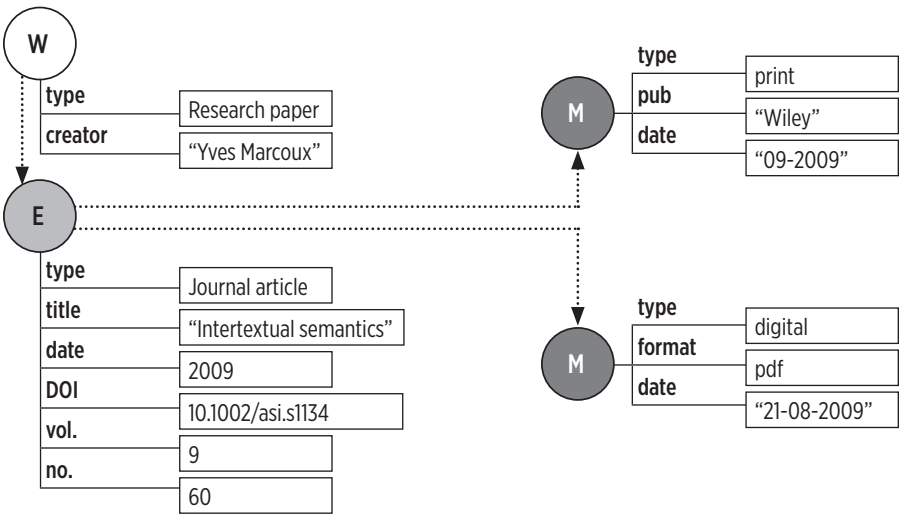
FaBiO FRBR-Aligned Bibliographic Ontology

FaBiO is one of the vocabularies defined in the Semantic Publishing and Referencing (SPAR) suite. This vocabulary development project has scholarly publishing as its main focus, with coverage primarily aimed at works that are published, textual, and/or referred to by bibliographic references. It defines its own set of entities that are sub-classed to the original FRBR entities (work, expression, etc.) using the FRBRcore RDF vocabulary. FaBiO's main approach to the FRBR entities is to use them as classes for types of scholarly publications. FaBiO sub-classes the FRBRcore work class with almost thirty different subtypes, including biography, reference work, dataset, and sound record. The FRBRcore expression class yields over fifty sub-classes, among which are chapter, editorial, presentation, spreadsheet, and Gantt chart.

In keeping with its approach of creating sub-classes of the FRBR entities defined as classes in FRBRcore, a manifestation in FaBiO is defined as having three sub-classes: analog manifestation, digital manifestation, and manifestation collection. The model is illustrated in figure 10.4.

FIGURE 10.4

FaBiO model showing expression and manifestation examples



Along with the classes derived from FRBR entities, FaBiO has dozens of properties for bibliographic description, few of which would be considered exact equivalents of descriptive elements in library data. There are some common bibliographic properties like title, publisher, data of publication, pages, identifier, and language, but FaBiO also has an extensive list of dates relating to the workflow of academic publishing, such as "has submission date," "has embargo date," and "has date received." FaBiO does not, however, align use of bibliographic properties to specific sub-classes of FRBR's WEMI. Instead, the descriptive properties are associated with the super-class that represents the bibliographic resource as a whole.

In addition to the extensions of FRBR provided by FRBRcore, FaBiO creates direct links between works and manifestation, works and items, and expressions and items.

Where FaBiO most notably diverges from a library cataloging interpretation of the entities, though, is in its emphasis on the expression. Like the expression-dominant model of Shoichi Taniguchi, FaBiO interprets the manifestation as a physical carrier, with all of the content description properties being associated with the expression. This model defines different digital formats of the same publication or manuscript in the item entity, which is a possible resolution to the

situation that libraries face with electronic books and other electronic resources where only the digital format varies.

Summary

FRBRer, FRBRcore, FaBiO, <indecs>, BIBFRAME, and RDA in RDF are variations on a theme. They have much in common, but each has its own approach. webFRBRer defines a strict closed-world interpretation of the FRBR model. RDA in RDF implements the elements of the FRBR-based cataloging rules. FRBRoo is highly sophisticated in its design but is so far unconnected to mainstream library cataloging. FRBRcore and BIBFRAME show promise as vocabularies for the open web, and FRBRcore has been the basis for the development of other bibliographic vocabularies. BIBFRAME, however, is taking on the unenviable task of carrying forward into RDF the centuries-old practices of traditional library cataloging and in particular that tradition as a coded machine-readable record, MARC 21.

None of these can be considered mature at this time, and the future of bibliographic data in RDF is still at an experimental stage. Although there is some speculation in the library community about which of these models will prevail, we have to consider the possibility that there will be more than one model in use. Already there are efforts to assure that bibliographic models under development promote interoperability, not only among the library, archives, and museum bibliographic models, but also with the worlds of publishing, academe, and the reading public.

AFTERWORD

There is no question that FRBR represents a great leap forward in the theory of bibliographic description. It addresses the "work question" that so troubled some of the great minds of library cataloging in the twentieth century. It provides a view of the "bibliographic family" through its recognition of the importance of the relationships that exist between created cultural objects. It has already resulted in vocabularies that make it possible to discuss the complex nature of the resources that libraries and archives gather and manage.

As a conceptual model, FRBR has informed a new era of library cataloging rules. It has been integrated into the cataloging workflow to a certain extent. FRBR has also inspired some nonlibrary efforts, and those have given us interesting insight into the potential of the conceptual model to support a variety of different needs.

The FRBR model, with its emphasis on bibliographic relationships, has the potential to restore context that was once managed through alphabetical collocation to the catalog. In fact, the use of a Semantic Web technology with a model of entities and relations could be a substantial improvement in this area, because the context that brings bibliographic units together can be made explicit: "translation of," "film adaptation of," "commentary on." This, of course, could be achieved with or without FRBR, but because the conceptual model articulates the relationships, and the relationships are included in the recent cataloging rules, it makes sense to begin with FRBR and evolve from there.

However, the gap between the goals developed at the Stockholm meeting in 1991 and the result of the FRBR Study Group's analysis is striking. FRBR defined only a small set of functional requirements, at a very broad level: find, identify, select, and obtain. The study would have been more convincing as a functional analysis if those four tasks had been further analyzed and had been the focus of the primary content of the study report. Instead, from my reading of the FRBR Final Report, it appears that the entity-relation analysis of bibliographic data took precedence over user tasks in the work of the FRBR Study Group.

The report's emphasis on the entity-relation model, and the inclusion of three simple diagrams in the report, is mostly likely the reason for the widespread belief that the FRBR Final Report defines a technology standard for bibliographic data. Although technology solutions can and have been developed around the FRBR conceptual model, no technology solution is presented in the FRBR Final Report. Even more importantly, there is nothing in the FRBR Final Report to suggest that there is one, and only one, technology possible based on the FRBR concepts. This is borne out by the examples we have of FRBR-based data models, each of which interprets the FRBR concepts to serve their particular set of needs. The strength of FRBR as a conceptual model is that it can support a variety of interpretations. FRBR can be a useful model for future developments, but it is a starting point, not a finalized product.

There is, of course, a need for technology standards that can be used to convey information about bibliographic resources. I say "standards" in the plural, because it is undeniable that the characteristics of libraries and their users have such a wide range of functions and needs that no one solution could possibly serve all. Well-designed standards create a minimum level of compliance that allows interoperability while permitting necessary variation to take place. A good example of this is the light bulb: with a defined standard base for the light bulb we have been able to move from incandescent to fluorescent and now to LED bulbs, all the time keeping our same lighting fixtures. We must do the same for bibliographic data so that we can address the need for variation in the different approaches between books and non-books, and between the requirements of the library catalog versus the use of bibliographic data in a commercial model or in a publication workflow.

Standardization on a single over-arching bibliographic model is not a reasonable solution. Instead, we should ask: "what are the minimum necessary points of compliance that will make interoperability possible between these various uses and users?" Interoperability needs to take place around the information and meaning carried in the bibliographic description, not in the structure that carries the data. What must be allowed to vary in our case is the technology that carries that message, because it is the rapid rate of technology change that we must be able to adjust to in the least disruptive way possible. The value of a strong conceptual model is that it is not dependent on any single technology.

It is now nearly twenty years since the Final Report of the FRBR Study Group was published. The FRBR concept has been expanded to include related standards for subjects and for persons, corporate bodies, and families. There is an ongoing Working Group for Functional Requirements for Bibliographic Records

that is part of the Cataloguing Section of the International Federation of Library Associations. It is taken for granted by many that future library systems will carry data organized around the FRBR groups of entities. I hope that the analysis that I have provided here encourages critical thinking about some of our assumptions, and fosters the kind of dialog that is needed for us to move fruitfully from broad concepts to an integrative approach for bibliographic data.

BIBLIOGRAPHY

ALA Committee on Co-Operation. "Standard Catalogue Card." *Library Journal* 1, no. 8 (1877): 285.

Albertsen, Ketil, and Carol van Nuys. "Paradigma: FRBR and Digital Documents." *Cataloging and Classification Quarterly* 39, no. 3–4 (2005): 37–41. doi:10.1300/J104v39n03_09.

Andersen, Jack. "Materiality of Works: The Bibliographic Record as Text." *Cataloging and Classification Quarterly* 33, no. 3/4 (November 2013): 39–65. doi:10.1300/J104v33n03_04.

Antelman, Kristin. "Identifying the Serial Work as a Bibliographic Entity" *Library Resources and Technical Services* 48, no. 4 (October 2004): 238–55.

Armstrong, Sharon Lee, Lila R. Gleitman, and Henry Gleitman. "What Some Concepts Might Not Be." In *Concepts: Core Readings,* edited by Eric Margolis and Stephen Laurence, 225–260. Cambridge, MA: MIT Press, 1999.

Avram, Henriette. *The MARC Pilot Project. Final Report.* Washington, DC: Library of Congress, 1968.

Baca, Murtha, 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,* edited by Arlene Taylor, 103–10. Westport, CT: Libraries Unlimited, 2007.

Baker, Thomas, Karen Coyle, and Sean Petiya. "Multi-Entity Models of Resource Description in the Semantic Web." *Library Hi Tech* 32, no. 4 (2014): 562–82. doi:10.1108/LHT-08–2014-0081.

Bean, Carol. *Relationships in the Organization of Knowledge.* Dordrecht; Boston; Norwell MA: Kluwer Academic Publishers, 2001.

Bekiari, Chryssoula, Martin Doerr, and Patrick Le Boeuf. *FRBR: Object-Oriented Definition and Mapping to FRBR(ER), V.1.0.2.* International Working Group on FRBR and CIDOC CRM Harmonisation, 2012.

Bennett, Rick, Brian F. Lavoie, and Edward T. O'Neill. "The Concept of a Work in WorldCat: An Application of FRBR." *Library Collections, Acquisitions, and Technical Services* 27, no. 1 (March 2003): 45–59. doi:10.1016/S1464-9055(02)00306-8.

Berg, Einar Silset. "A Comparison of Two Relational Models: IFLA's FRBR Model and Taniguchi's Expression-Prioritized Model." Master's thesis, Oslo University, 2004.

Berners-Lee, Tim, and Ora Lassila. "The Semantic Web." *Scientific American* (May 2001). http://www.scientificamerican.com/article.cfm?id=the-semantic-web.

Bianchini, Carlo, and Mauro Guerrini. "From Bibliographic Models to Cataloging Rules: Remarks on FRBR, ICP, ISBD, and RDA and the Relationships between Them." *Cataloging and Classification Quarterly* 47, no. 2 (January 2009): 105–24. doi:10.1080/01639370802561674.

Buckland, Michael K. "What Is a 'Document'?" *Journal of the American Society for Information Science* 48, no. 9 (1997): 804–9.

Calhoun, Karen. *The Changing Nature of the Catalog and Its Integration with Other Discovery Tools*. Library of Congress, 2006. http://www.loc.gov/catdir/calhoun-report-final.pdf.

Carlyle, Allyson. "Understanding FRBR as a Conceptual Model: FRBR and the Bibliographic Universe." *Library Resources and Technical Services* 50, no. 4 (2006): 264–73.

Carpenter, Michael, and Elaine Svenonius, eds. *Foundations of Descriptive Cataloging* Littleton, CO: Libraries Unlimited, 1985.

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

Chen, Peter Pin-shan. "The Entity-Relationship Model—Toward a Unified View of Data." *ACM Transactions on Database Systems* 1, no. 1 (1976): 9–36.

Cossham, Amanda F. "Bibliographic Records in an Online Environment." *IR Information Research* 18, no. 3 (September 2013). http://www.informationr.net/ir/18–3/colis/paperC42.html.

Coyle, Karen. "FRBR, the Domain Model." *Library Technology Reports* 46, no. 2 (2010): 20–25.

____. "FRBR, Twenty Years On." *Cataloging and Classification Quarterly* 32, no. 4 (December 2014): 1–21. doi:10.1080/01639374.2014.943446.

Cutter, Charles Ammi. "The New Catalogue of the Harvard College Library." *North American Review* (January 1869).

____. *Rules for a Dictionary Catalog*. Washington, DC: US Bureau of Education, 1904.

____. *Rules for a Dictionary Catalogue*. Washington, DC: Library of Congress, 1875.

Davis, Ian, and Richard Newman. *Expression of Core FRBR Concepts in RDF*. 2005. http://vocab.org/frbr/core.html.

Delsey, Tom. *RDA Database Implementation Scenarios,* 2009. http://www.rda-jsc.org/docs/5editor2rev.pdf.

____. "Standards for Descriptive Cataloging: Two Perspectives on the Past Twenty Years." In *The Conceptual Foundations of Descriptive Cataloging,* edited by Elaine Svenonius, 51–60. San Diego, CA: Academic Press, 1989.

Denton, William. "FRBR and the History of Cataloging." In *Understanding FRBR,* 35–57, n.d.

____. "My Speaking Notes for 'What We Talk About When We Talk About FRBR,'" Presentation at Code4Lib 2009" 2009, Providence, RI, February 2009. http://yorkspace.library.yorku.ca/xmlui/bitstream/handle/10315/2528/frbr_code4lib09_denton_notes.pdf.

Dervin, B. "The Everyday Information Needs of the Average Citizen: A Taxonomy for Analysis." In *Information for the Community,* edited by Manfred Kochen and Joseph C. Donohue. Chicago: American Library Association, 1976.

Dickey, Timothy J. "FRBRization of a Library Catalog: Better Collocation of Records, Leading to Enhanced Search, Retrieval, and Display." *Information Technology and Libraries* 27 (2008): 23–32. doi:10.6017/ital.v27i1.3260.

Doerr, Martin, and Patrick LeBoeuf. "Modelling Intellectual Processes: The FRBR-CRM Harmonization" In *Digital Libraries: Research and Development. First International DELOS Conference, Pisa, Italy, February 13-14, 2007.* Berlin Heidelberg, Springer-Verlag, 2007. doi:10.1007/978-3-540-77088-6_11.

Doerr, Martin, Guillaume Boutard, Günther Görz, Dolores Iorizzo, Max Jacob, Carlos Lamsfus, Mika Nyman, et al. *FRBR Object-Oriented Definition and Mapping to FRBRer,* 2007.

Domanovszky, Ákos. *Functions and Objects of Author and Title Cataloguing: A Contribution to Cataloguing Theory.* München: Verlag Dokumentation. 1975.

Dunsire, Gordon. "Representing the FR Family in the Semantic Web." *Cataloging and Classification Quarterly* 50, no. 5–7 (June 2012): 724–41. doi:10.1080/01639374.2012.679881.

____. *Resource and Work, Expression, Manifestation, Item,* 2013. http://www.ifla.org/files/assets/cataloguing/isbd/OtherDocumentation/resource-wemi.pdf.

Garrison, Dee. *Apostles of Culture: The Public Librarian and American Society, 1876–1920.* New York: Macmillan Information, 1979.

Genette, Gerard. *Paratexts: Thresholds of Interpretation*. Cambridge [Eng.]: Cambridge University Press, 1997.

Gorman, Michael. "AACR3? Not!" *ALCTS Papers on Library Technical Services and Collections* (1998): 19–29.

Gorman, Michael, and Paul W. Winkler, eds. "25.1 Use of Uniform Titles." In *Anglo-American Cataloguing Rules; Second Edition,* 441–42. Chicago: American Library Association, 1978.

Gorman, Michael, Paul W. Winkler, and American Library Association. *Anglo-American Cataloguing Rules.* 2nd ed. London: Library Association, 1978.

Hardesty, Juliet L., Steven Harris, Anna Coogan, and Mark Notess. *Scherzo Usability Test Report: Testing a FRBR Search Interface for Music.* Indiana University; Variations project, 2012.

Hickey, Thomas B., and Edward T. O'Neill. "Experiments with the IFLA Functional Requirements for Bibliographic Records (FRBR)," 2002, 1–15.

Hickey, Thomas B., and Jenny Toves. "FRBR Work-Set Algorithm." Dublin, OH: OCLC, 2005. http://www.oclc.org/content/dam/research/activities/frbralgorithm/2005-04.pdf.

Hillmann, Diane, Karen Coyle, John Phipps, and Gordon Dunsire. "RDA Vocabularies: Process, Outcome, Use." *D-Lib Magazine* 16, no. 1/2 (2010). doi:10.1045/january2010-hillmann.

Howarth, Lynne C. "FRBR and Linked Data: Connecting FRBR and Linked Data." *Cataloging and Classification Quarterly* 50, no. 5–7 (June 2012): 763–76. doi:10.1080/01639374.2012.680835.

IFLA (International Federation of Library Associations). *Statement of International Cataloging Principles,* (2009).

———. *Statement of Principles Adopted by the International Conference on Cataloguing Principles.* Paris, 1961.

IFLA Cataloguing Section. *International Standard Bibliographic Description (ISBD),* 2011. http://www.ifla.org/files/assets/cataloguing/isbd/isbd-cons_20110321.pdf.

IFLA Cataloguing Section. ISBD Review Group. *Mapping ISBD Elements to FRBR Entity Attributes and Relationships.* IFLA, 2004. http://www.ifla.org/files/assets/cataloguing/isbd/isbd-frbr-mapping.pdf.

IFLA Study Group on the Functional Requirements for Bibliographic Records. *Functional Requirements for Bibliographic Records,* 2009. http://archive.ifla.org/VII/s13/frbr/frbr_2008.pdf.

IFLA Standing Committee of the Section on Cataloguing, Terms of Reference for a Study of the Functional Requirements for Bibliographic Records, 1992.

International Working Group on FRBR and CIDOC CRM Harmonisation. *FRBR Object-Oriented Definition and Mapping from FRBR$_{ER}$, FRAD and FRSAD.* CIDOC, 2015. http://www.cidoc-crm.org/docs/frbr_oo//frbr_docs/FRBRoo_V2.1_2015February.pdf.

Jones, Ed. "The FRBR Model as Applied to Continuing Resources." *Library Resources and Technical Services* 49, no. 4 (October 2005.): 227–42.

Jones, Ed, Caitlin Hunter, and Jenn Riley. *FRBR, FRAD and Music: Theory and Practice* Presentation, Music Library Association, 2010. http://bcc.musiclibraryassoc.org/Presentations/FRBR-FRAD-MLA-2010.pdf.

Jonsson, Gunilla. "Cataloguing of Hand Press Materials and the Concept of Expression in FRBR," *Cataloging and Classification Quarterly,* November 2013 (2009): 37–41. doi:10.1300/J104v39n03.

Kalf, Ruth. "FRBR: An Opportunity for Map Collections and Map Users?" *LIBER Quarterly* 18, no. 2 (September 2008). https://liber.library.uu.nl/index.php/lq/article/viewFile/7927/8185.

Knapp, John F. "Design Considerations for the MARC Magnetic Tape Formats." *Library Resources and Technical Services* 12, no. 3 (1968): 275–85.

Kochen, Manfred, and Joseph C. Donohue, eds. *Information for the Community.* Chicago: American Library Association, 1976.

Krajewski, Markus. *Paper Machines: About Cards and Catalogs, 1548–1929.* Cambridge, MA: MIT Press, 2011.

Krier, Laura. "Serials, FRBR, and Library Linked Data: A Way Forward." *Journal of Library Metadata* 12, no. 2–3 (April 2012): 177–87. doi:10.1080/19386389.2012.699834.

Lavoie, Brian F., and Edward T. O'Neill. "The Concept of a Work in WorldCat: An Application of FRBR." *Library Collections, Acquisitions, and Technical Services* 27, no. 1 (2003): 32.

Le Boeuf, Patrick. "Foreword." *Cataloging and Classification Quarterly* 50, no. 5–7 (June 2012): 355–59. doi:10.1080/01639374.2012.682001.

____. "FRBR: Hype or Cure-All? Introduction." In *Functional Requirements for Bibliographic Records: Hype or Cure-All?* 1–13. New York: Haworth Information Press, 2005.

____. *Functional Requirements for Bibliographic Records (FRBR): Hype or Cure-All?* Edited by Patrick Le Boeuf. New York: Haworth Information Press, 2005.

Le Boeuf, Patrick, Barbara Tillett, and Elena Escolano Rodriguez. *Brave New FRBR World*. Cairo: IME ICC 3, 2005.

Lerner, Frederick Andrew. *The Story of Libraries: From the Invention of Writing to the Computer Age*. New York: Continuum, 2001.

Library of Congress Working Group on the Future of Bibliographic Control. *On the Record*. Washington, DC: Library of Congress, 2008.

Lubetzky, Seymour. *Seymour Lubtezky: Writings on the Classical Art of Cataloging*. Edited by Elaine Svenonius and Dorothy McGarry. Englewood, CO, Libraries Unlimited, 2001.

Lugg, Rick, and R2 Consulting LLC. "Working Group on the Future of Bibliographic Control: Economics and Organization of Bibliographic Data Economics of Bibliographic Control," 2007. http://www.loc.gov/bibliographic-future/meetings/docs/ricklugg-july9-2007.pdf.

Madison, Olivia M. A. "The IFLA Functional Requirements for Bibliographic Records: International Standards for Universal Bibliographic Control." *Library Resources and Technical Services* 44, no. 3 (July 2000): 153–159.

____. "The Origins of the IFLA Study on Functional Requirements for Bibliographic Records." *Cataloging and Classification Quarterly* 39, no. 3–4 (2005): 15–37.

____. "Standards for International Bibliographic Control: Proposed Basic Data Requirements for the National Bibliographic Record." In *Universal Bibliographic Control and International MARC Core Programme (UBCIM),* 1–11. Copenhagen: ICNBS, 1998.

Malssen, Kara Van, Caitlin Hunter, and Andrea Leigh. *BIBFRAME AV Modeling Study : Defining a Flexible Model for Description of Audiovisual Resources*. Library of Congress, 2014. http://www.loc.gov/bibframe/pdf/bibframe-avmodelingstudy-may15-2014.pdf.

Manzanos, Norberto. "Item, Document, Carrier: An Object Oriented Approach." *Cataloging and Classification Quarterly* 50, no. February 2014 (2012): 439–55. doi: 10.1080/01639374.2012.681274.

Marcum, Deanna B. "Response to *On the Record : Report of the Library of Congress Working Group on the Future of Bibliographic Control*." Washington, DC: Library of Congress. 2008. http://www.loc.gov/bibliographic-future/news/LCWGResponse-Marcum-Final-061008.pdf.

Margolis, Eric, and Stephen Laurence. *Concepts: Core Reading*. Cambridge, MA: MIT Press, 1999.

Maunsell, Andrew. *The Catalogue of English Printed Books*. Farnborough, UK: Gregg, Press, 1965. First published 1595.

Maxwell, Robert L. *FRBR: A Guide for the Perplexed*. Chicago: American Library Association, 2008.

Miksa, Shawne D. "Understanding Support of FRBR's Four User Tasks in MARC-Encoded Bibliographic Records." *Bulletin of the American Society for Information Science and Technology,* August/September 2007: 24–26.

Miller, David, and Patrick Le Boeuf. "'Such Stuff as Dreams Are Made On': How Does FRBR Fit Performing Arts ?" *Cataloging and Classification Quarterly* 39, no. 3–4 (2005): 151–78. doi:10.1300/J104v39n03_10.

Murphy, Gregory L. *The Big Book of Concepts.* Cambridge, MA: MIT Press, 2004.

Murray, Ronald J. *The FRBR-Theoretic Library : The Role of Conceptual Data Modeling in Cultural Heritage Information System Design.* Washington, DC: Library of Congress, 2008. http://www.bl.uk/ipres2008/presentations_day1/26_Murray.pdf.

Murray, Ronald J., and Barbara B. Tillett. "From Moby-Dick to Mashups : Thinking about Bibliographic Networks Where This Is Going :" In *American Library Association 2010 Annual Conference.* Washington, DC: Library of Congress, 2010. http://www.slideshare.net/RonMurray/from-mobydick-to-mashups.

Noerr, Peter, Paula Goossens, Dan Matei, Petra Otten, and Susanna Peruginelli. "User Benefits from a New Bibliographic Model: Follow-up of the IFLA Functional Requirements Study." In *64th IFLA General Conference,* 1–6. IFLA, 1998. http://archive.ifla.org/IV/ifla64/084-126e.htm.

O'Neill, Ed, Anders Cato, Paula Goossens, Judy Kuhagen, Barbara Tillett, Carol Van Nuys, and Maja Žumer. *Final Report of the Working Group on Aggregates,* 2011. http://www.ifla.org/files/assets/cataloguing/frbrrg/AggregatesFinalReport.pdf.

O'Neill, Edward T. "FRBR : Functional Requirements for Bibliographic Records Application of the Entity-Relationship Model to Humphry Clinker." *Library Resources and Technical Services* 46, no. 4 (2002): 150–59.

O'Neill, Edward T., and Maja Žumer. *Aggregates as Manifestations.* IFLA Working Group on Aggregates, n.d.

Pearl, Nancy. *Book Lust: Recommended Reading for Every Mood, Moment, and Reason.* Seattle: Sasquatch Books. 2003.

Peponakis, Manolis. "Conceptualizations of the Cataloging Object: A Critique on Current Perceptions of FRBR Group 1 Entities." *Cataloging and Classification Quarterly* 50, no. 5–7 (June 2012): 587–602. doi:10.1080/01639374.2012.681275.

____. "Libraries' Metadata as Data in the Era of the Semantic Web Manolis Peponakis." *Journal of Library Metadata* 13, no. 4 (2013): 330–48. doi:10.1080/19386389.2013.846618.

Perkinson, Richard C. *Data Analysis: The Key to Data Base Design.* Wellesley, MA: QED Information Sciences, 1984.

Petrucciani, Alberto. "From the FRBR Model to the Italian Cataloging Code (and Vice Versa?)." *Cataloging and Classification Quarterly* 50, no. 5–7 (June 2012): 603–21. doi:10.1080/01639374.2012.680841.

Picco, Paola, and Virginia Ortiz Repiso. "The Contribution of FRBR to the Identification of Bibliographic Relationships: The New RDA-Based Ways of Representing Relationships in Catalogs." *Cataloging and Classification Quarterly* 50, no. 5–7 (June 2012): 622–40. doi:10.1080/01639374.2012.680847.

Pisanski, Jan, and Maja Žumer. "Functional Requirements for Bibliographic Records: An Investigation of Two Prototypes." *Program: Electronic Library and Information Systems* 41, no. 4 (2007): 400–417. doi:10.1108/00330330710831602.

____. "Mental Models of the Bibliographic Universe. Part 1: Mental Models of Descriptions." *Journal of Documentation* 66, no. 5 (2010): 643–67.

____. "Mental Models of the Bibliographic Universe. Part 2: Comparison Task and Conclusions." *Journal of Documentation* 66, no. 5 (2010): 668–80.

Ranganathan, S. R. *The Five Laws of Library Science*. Bombay, New York: Asia Pub. House, 1963.

Ranganathan, S. R., and M. A. Gopinath. *Prolegomena to Library Classification*. New Delhi: Ess Ess Publications, 1989.

Rather, Lucia J. "Special Characters and Diacritical Marks Used in Roman Alphabets." *Library Resources and Technical Services* 12, no. 3 (1968): 285–95.

Renear, Allen, and Yunseon Choi. "Some Conceptual Modeling Issues in FRBR," 2006, 1–7.

Renear, Allen H., and David Dubin. "Three of the Four FRBR Group 1 Entity Types Are Roles, Not Types." *Proceedings of the American Society for Information Science and Technology* 44, no. 1 (October 2008): 1–19. doi:10.1002/meet.1450440248.

Renear, Allen H., and Yunseon Choi. "Modeling Our Understanding, Understanding Our Models—The Case of Inheritance in FRBR." *Proceedings of the American Society for Information Science and Technology.* New York: Wiley, 2006 doi:10.1002/meet.14504301179.

Riley, Jenn. "Application of the Functional Requirements for Bibliographic Records (FRBR) to Music." In *Proceedings of the International Society for Music Information Retrieval 2008,* 439–44, 2008.

____. "Enhancing Interoperability of FRBR-Based Metadata." In *International Conference on Dublin Core and Metadata Applications,* 31–43, 2010.

Riley, Jenn, Caitlin Hunter, Chris Colvard, and Alex Berry. *Definition of a FRBR-Based Metadata Model for the Indiana University Variations3 Project,* 2007. http://dlib.indiana.edu/projects/variations3/docs/v3FRBRreport.pdf.

Riva, Pat, and Chris Oliver. "Evaluation of RDA as an Implementation of FRBR and FRAD." *Cataloging and Classification Quarterly* 50, no. 5–7 (June 2012): 564–86. doi:10.1080/01639374.2012.680848.

Rose, Mary Z. "The Ship Has Sailed and We Aren't on It: How Catalogers Could Support User Tasks and Why We Won't." *Journal of Library Metadata* 12, no. 2–3 (April 2012): 127–39. doi:10.1080/19386389.2012.699828.

Rust, Godfrey, and Mark Bide. *The Indecs Metadata Framework; Principles, Model and Data Dictionary*. Indecs Framework Ltd, 2000.

Schaub, Jacob, Yin Zhang, and Athena Salaba. "What Is Next for Functional Requirements for Bibliographic Records? A Delphi Study." *Library Quarterly* 79, no. 2 (2009): 233–55.

Schmidt, Raymond. "Composing in Real Time: Jazz Performances as 'Works' in the FRBR Model." *Cataloging and Classification Quarterly* 50, no. 5–7 (June 2012): 653–69. doi:10.1080/01639374.2012.681601.

Seikel, Michele. "General Notes in Catalog Records versus FRBR User Tasks." *Cataloging and Classification Quarterly* 51, no. 4 (May 2013): 420–27. doi:10.1080/01639374.2012.749318.

Serrai, Alfredo, and Fiammetta Sabba. 2005. *Profilo di storia della bibliografia*. Milano: Sylvestre Bonnard.

Shadle, Steve. "FRBR and Serials: An Overview and Analysis." *Serials Librarian* 50, no. 1–2 (2008): 83–103. doi:10.1300/J123v50n01_09.

Shotton, David. *Extending FRBR within FaBiO,* 2011. http://opencitations.wordpress.com/2011/06/29/extending-frbr-within-fabio/.

Smiraglia, Richard P. "Be Careful What You Wish For: FRBR, Some Lacunae, a Review." *Cataloging and Classification Quarterly* 50, no. 5–7 (June 2012): 360–68. doi:10.1080/01639374.2012.682254.

____. "Further Reflections on the Nature of 'A Work': An Introduction." *Cataloging and Classification Quarterly* 33, no. 3–4 (August 2002): 1–11. doi:10.1300/J104v33n03_01.

____. "Instantiation: Toward a Theory." In *Data, Information and Knowledge in a Networked World; Annual Conference of the Canadian Association for Information Science. London, Ontario. June 2–4 2005,* 1–8, 2005.

____. "A Meta-Analysis of Instantiation as a Phenomenon of Information Objects." *Culture Del Testo e Del Documento* 9, no. 25 (2008): 5–25.

____. "Musical Works as Information Retrieval Entities: Epistemological Perspectives," 2nd Annual International Symposium on Music Information Retrieval, Bloomington, 2001. http://ismir2001.ismir.net/pdf/smiraglia.pdf.

____. *The Nature of "A Work": Implications for the Organization of Knowledge*. Lanham: Scarecrow Press, 2001.

____. *Works as Entities for Information Retrieval.* Binghampton, NY: Haworth Information Press. 2002.

Spalding, Tim. IMPORTANT: LibraryThing Dives into Editions and Expressions, 2011. http://www.librarything.com/topic/109523.

Svenonius, Elaine. *The Conceptual Foundations of Descriptive Cataloging.* San Diego, CA: Academic Press, 1989.

Taniguchi, Shoichi. "A Conceptual Model Giving Primacy to Text-Level Bibliographic Entity in Cataloging: A Detailed Discussion," *Journal of Documentation* 58, no. 4 (4) (2003): 363–82.

____. "Expression-Level Bibliographic Entity Records: A Trial on Creation from Pre-Existing MARC Records," (October 2008): 37–41. doi:10.1300/J104v38n02_05.

____. "User Tasks in the RDA-Based Model." *Cataloging and Classification Quarterly* 51, no. 7 (October 2013): 788–815. doi:10.1080/01639374.2013.800623.

____. "Viewing RDA from FRBR and FRAD: Does RDA Represent a Different Conceptual Model?" *Cataloging and Classification Quarterly* 50, no. 8 (November 2012): 929–43. doi:10.1080/01639374.2012.712631.

Taylor, Arlene, ed. *Understanding FRBR: What It Is and How It Will Affect Our Retrieval Tools.* Westport, CT: Libraries Unlimited, 2007.

Thomale, Jason. "Interpreting MARC: Where's the Bibliographic Data?" *Code4lib Journal,* no. 11 (September 21, 2010).

Tillett, Barbara B. "AACR3: Resource Description and Access." Presentation to Library of Congress, 2004.

____. "Bibliographic Relationships." In *Relationships in the Organization of Knowledge,* 19–36. Dordrecht: Kluwer, 2001.

____. "Bibliographic Relationships: Toward a Conceptual Structure of Bibliographic information Used in Cataloging." Dissertation, University of California, Los Angeles, 1988.

____. *Definition of Aggregates as Works: Tillett Proposal,* IFLA Working Group on Aggregates, 2007.

____. "FRBR and Cataloging for the Future." *Cataloging and Classification Quarterly* 39, no. 3 (2005): 197–205.

____. "IFLA Study on the Functional Requirements of Bibliographic Records: Theoretical and Practical Foundations," Proceedings of the 60th IFLA General Conference, August 21–27, 1994.

____. *What Is FRBR? A Conceptual Model for the Bibliographic Universe.* Washington, DC: Library of Congress. Cataloging Distribution Service, 2003.

Turner, James M, and Abby A Goodrum. "Modeling Videos as Works." *Cataloging and Classification Quarterly* 33, no. 3–4 (2009): 27–38. doi:10.1300/J104v33n03_03.

University of Minnesota Libraries. *A Multi-Dimensional Framework for Academic Support: A Final Report,* 2006.

Unsworth, John. "Scholarly Primitives: What Methods Do Humanities Researchers Have in Common and How Might Our Tools Reflect This?" In *Humanities Computing, Formal Methods, Experimental Practice,* Kings College, London, May 13, 2000, 2000.

Verona, Eva. "Literary Unit versus Bibliographic Unit." In *Foundations of Descriptive Cataloging,* ed. By Michael Carpenter and Elaine Svenonius, 155–175. Littleton, CO: Libraries Unlimited, 1985.

Weiss, Paul J., and Steve Shadle. "FRBR in the Real World." *Serials Librarian* 52, no. 1–2 (2007): 93–104. doi: 10.1300/J123v52n01_09.

Wiesenmüller, Heidrun. *Some Comments on the Final Report of the FRBR Working Group on Aggregates.* Stuttgart: Stuttgart Media University, 2011. https://www.mendeley.com/download/public/5179751/4486442993/59a97ce9df66ece320e2dc2e5e6421730c0531d4/dl.pdf.

Willer, Mirna, and Ana Barbari. "FRBR/FRAD and Eva Verona's Cataloging Code: Toward the Future Development of Croatian Cataloging Code." *Cataloging and Classification Quarterly* 50, no. 5–7 (June 2012): 542–63. doi:10.1080/01639374.2012.681613.

Wilson, Patrick. "Interpreting the Second Objective of the Catalog." *Library Quarterly* 59, no. 4 (1989): 339–53.

____. "The Second Objective." In *The Conceptual Foundations of Descriptive Cataloging,* edited by Elaine Svenonius, 5–16. San Diego, CA: Academic Press, 1989.

____. *Two Kinds of Power: An Essay on Bibliographical Control.* University of California Publications in Librarianship. Berkeley, CA: University of California Press, 1968.

Wilson, T. D. "Information-Seeking Behaviour and Information Use." *Social Science Information Studies* 4 (1984): 197–204.

____. "Models in Information Behavior Research." *Journal of Documentation* 22, no. 3 (1966): 266–68. doi:10.1108/eb026404.

Yee, Martha M.*Cataloging Rules and the RDF Model.* 2007. http://myee.bol.ucla.edu/catrul&rdfIntro.htm.

____. "FRBR and Moving Image Materials: Content (Work and Expression) Versus Carrier (Manifestation)." In *Understanding FRBR: What It Is and How It Will Affect Our Retrieval Tools,* edited by Arlene Taylor, 117–30. Westport, CT: Libraries Unlimited, 2007.

____. "FRBRization: A Method for Turning Online Public Finding Lists into Online Public Catalogs." *Information Technology and Libraries* 24, no. 3 (2005): 77–95.

Zhang, Yin, and Athena Salaba. "What Do Users Tell Us About FRBR-Based Catalogs?" *Cataloging and Classification Quarterly* 50, no. 5–7 (June 2012): 705–23. doi:10.1080/01639374.2012.682000.

Zipf, George Kingsley. *Human Behavior and the Principle of Least Effort: An Introduction to Human Ecology*. Cambridge, MA: Addison-Wesley Press. 1949.

INDEX