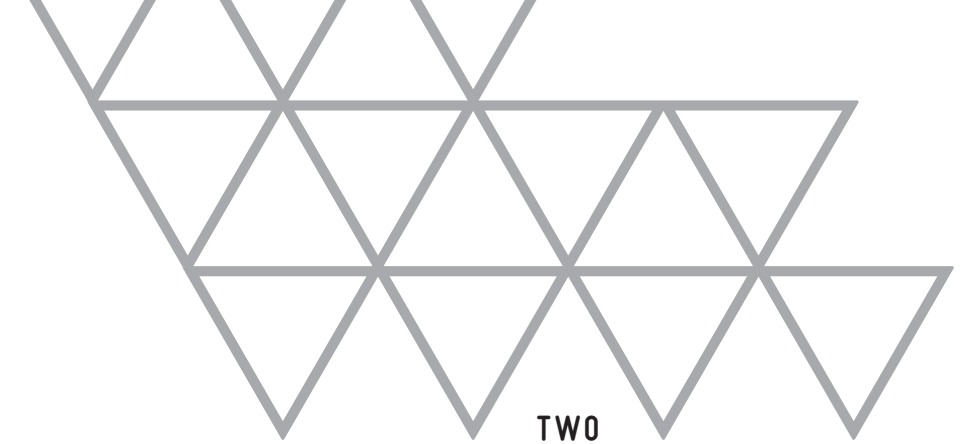

TWO

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.