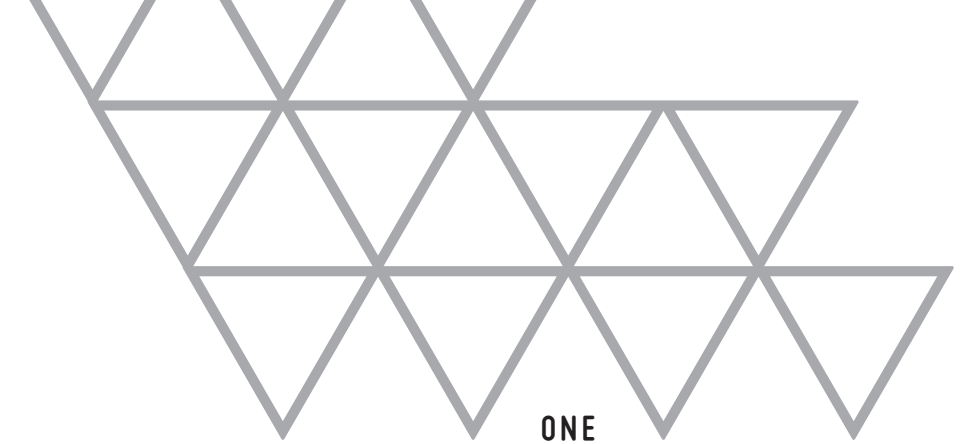


ONE

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)

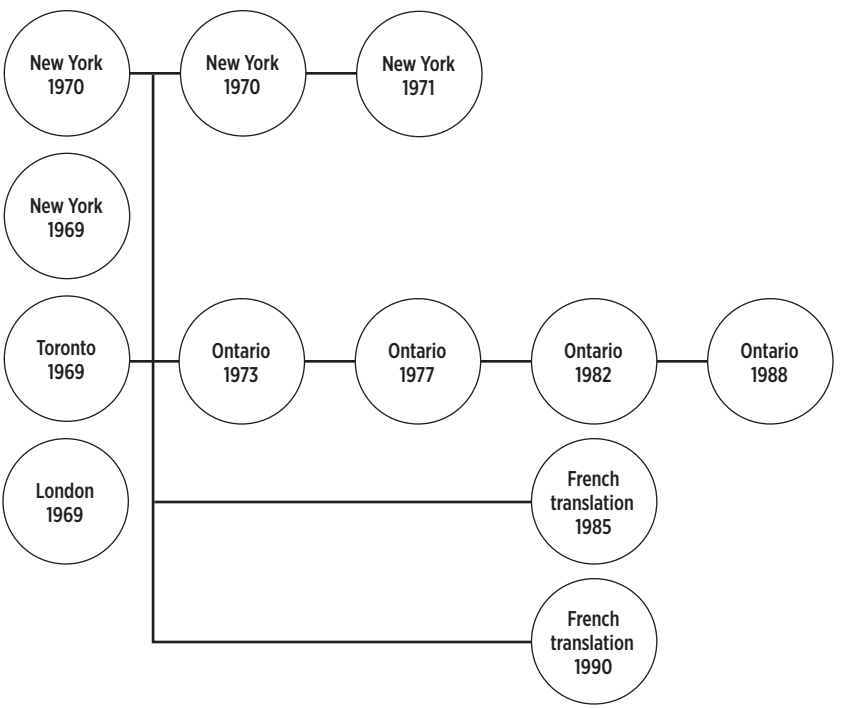
Wilson's view is one possible interpretation of S. R. Ranganathan's statement that "a library is a growing organism." In Wilson's view, the library grows not only in the number of volumes, but with the addition of volumes families grow in a variety of ways. Each addition to the library potentially adds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