

Ubiquitous Cataloging

Bradley Dilger, English & Journalism

Bill Thompson, University Libraries

Western Illinois University

Cataloging becomes ubiquitous

Ask library users to describe their Online Public Access Catalog (OPAC), and then to describe Google, and most will give indistinguishable answers. And this is no surprise. In many libraries, the standalone, single-function card catalog, designed to find books on nearby shelves, has been all but replaced by integrated, multi-function catalogs which not only search local holdings, but can also search the holdings of other libraries—and can request and check out items from afar. However, library users' conflation of OPACs and other entities also exists because Amazon.com, Google Books, and similar sites are in fact the primary catalogs for many people. In fact, LibraryThing, a popular social network and book catalog, uses either Amazon.com or the Library of Congress as its two main default sources for gathering book information suggesting an equivalence between the two. Further blurring long-established boundaries, as more content becomes available online in alternative catalogs and other databases, catalogs are not only a pathway to texts, but a final destination. Users rely on OPACs not only to locate NetLibrary books or government documents, but to retrieve them as well—if they are not already bypassing local catalogs in favor of Google Books.

Most importantly, as we enter the dawning age of ubiquitous computing, what Adam Greenfield calls “everyware,” we are also entering the dawning age of ubiquitous cataloging. Indeed, we live in the golden age of catalogs, because cataloging, once the province of scholars armed with pen and paper, has become one of the core skills of computing—a “unit operation,” to use the framework proposed by Ian Bogost (2006, 3). More and more of our daily activities

involve cataloging, which we believe must be considered in five ways. The first is the traditional sense: categorization and/or classification by trained ontologists guided by established standards. Librarians all over the world participate in this work every day. The second is categorization and/or classification by people without professional training, likely using no formal guidelines. Numerous web sites considered part of “Web 2.0,” which revolves around network-based applications and user input, as opposed to more traditional web sites which broadcast information, facilitate this type of cataloging. LibraryThing and Amazon.com fit this bill, as do bookmark managers like del.icio.us and photography sharing sites such as Flickr. The third might be called “catalog-networking:” the act of browsing catalogs, comparing them, and connecting all kinds of information systems. In the age of Google, this activity happens continuously, at the hands of both professional ontologists, and those who catalog-network for fun, at school, or on the job (often through the same Web 2.0 sites). The fourth is the design of catalogs and cataloging interfaces, whether OPACs, less traditional systems like LibraryThing, or even search functionalities like Apple’s Spotlight for Mac OS X. Obviously, as information systems proliferate, so does the need to manage the information they contain and thus this manner of cataloging as well. The fifth and final type is cataloging of catalogers: the collection, storage, and retrieval of information about catalog users, in both contexts immediately connected to, and seemingly far from, traditional cataloging. Today a trip to the market can initiate this kind of cataloging: as customers swipe a key fob required for a discount, databases quietly plan restocking of inventories and accumulate information about purchases which can be articulated to consumer demographics. (In library terms, the grocery store “engages in collection development programs appropriate to its user community.”)

A brief case study of LibraryThing illustrates the importance and interconnectedness of these five senses of cataloging. The primary purpose of the site is building a personal catalog:

one's own searchable and sortable database of books, accessible from any device with a web browser. What if the patrons of a library could voluntarily make these off-site catalogs visible to other interested readers? Then library catalogs would create connections between patrons—user communities. LibraryThing is designed this way: it allows readers to share their lists of books they have purchased or read, to discover others who have read the same texts. Readers can import records from numerous sources of bibliographic information; as noted above, Amazon.com, the Library of Congress, and numerous other sources. Once they have imported a record, readers can add subject headings of their own, called “tags,” which can be matched against other LibraryThing users—providing another way to discover texts. Additionally, readers can quickly link via author or title to public records in any individual's personal collection. This double functionality enables readers to create a private catalog of their own reading, while simultaneously creating a “union catalog” linking all readers. The site also allows users to review books, comment on the reviews of other readers and facilitates intra-reader communication via email or online groups. In other words, site designer Tim Spalding and the community of LibraryThing users who have contributed to its development have covered all five senses of cataloging as we've defined it: integrating (1) established, formal cataloging protocols and catalogs with (2) user-contributed data, while (3) facilitating comparisons between wide varieties of formal and informal cataloging systems. Also, LibraryThing (4) constantly updates its design and functionality to meet user needs, while (5) allowing users to see each other while maintaining control over their own privacy.

LibraryThing and similar Web 2.0 sites provide models of cataloging to be carefully studied and emulated by all kinds of libraries (Miller 2005). The OPAC monopoly on cataloging is dead, thanks to Google, LibraryThing and a host of other web-based systems which facilitate one or more of our five kinds of cataloging. While OPACs are far from dead, unless they change,

library catalogs and the work catalogers will be downgraded to a finding aid, a quick stop in a circuit which begins with and returns to Amazon.com or Google Books. Indeed, Google's massive text digitization project is changing the way people locate information in and about books, and could do even more: as a recent *New Yorker* article estimated, digitizing all 32 million books in WorldCat would cost about \$800 million—well within the reach of a company capitalized at \$150 billion (Toobin 2007, 33). In this essay, we join those arguing for expanding the functions of library catalogs, and calling on cataloging vendors to do the same (Byrum 2005, Lossau 2004). In this way, libraries can maintain their commitment to user-centered technology while keeping the catalog centered in the library—no longer a given, as D. Grant Campbell and Karl V. Fast point out (Campbell and Fast 2004, 26). This work would simultaneously enhance libraries' position as the center of community information networks. Imagine a map which merged LibraryThing data and Google Maps to show the distribution of local readers of Agatha Christie mysteries; that would provide an entirely new kind of community analysis. Libraries can enable these new kinds of information resources while answering their core missions and continuing to defend the privacy rights of their patrons. To these ends, we draw upon network theory to show why and how librarians should consider cataloging broadly, as we do, and to outline methods for answering the implicit challenge of LibraryThing and other alternative catalogs emerging from Web 2.0—or rather, for seeing these information systems as possible partners, not competitors, in a world of ubiquitous cataloging.

Catalogs and networks

Most library patrons would agree library catalogs allow users to locate books and other materials by author, title, and subject. But most librarians would hasten to point out they can do much more. Catalogs allow connections between diverse subjects, provide pathways for researchers to contact each other, etc. To use the terms of network theory, catalogs are

simultaneously points of entry into networks, network nodes, and even means for making new nodes and connections. As we note above, we are not the first to apply network theory to library catalogs; they have always had a social component, enabling connections between readers and authors (Campbell and Fast 2004, 35). Traditionally, library catalogs create what Mark Granovetter called a “bridge,” the only path between one knowledge network (the library patron) and another (the particular work or works of interest in a library’s collections). The catalog allows users to discover, identify, and with the help of the shelf classification system, retrieve specific works new to them, bringing them into their knowledge networks. In this manner, catalogs aid both research, discovering knowledge, and findability, locating a particular object or idea (Morville 2005, 4). Arguably, connecting users and the works otherwise unknown to them remains the most critical function of library catalogs. But the connections catalogs facilitate aren’t always unique, especially for experienced researchers who are aware of many important authors and texts, even those not in their fields. Granovetter famously called these connections between two otherwise disconnected networks “weak ties.” His analogy compares acquaintances and friends (Granovetter 1973, 1366). With friends we have strong ties. We communicate with them routinely and share quite a bit of their knowledge, which is why so much can go unsaid between friends (Granovetter 1973, 1371–73). With acquaintances (weak ties), we communicate less frequently, perhaps rarely, but they know people and possess information our friends don’t. Weak ties are powerful because their weakness provides an entrée into the unknown, into new areas of knowledge where we can make discoveries and extend or even radically change our knowledge networks. While the information redundancy characteristic of strong ties allows close-knit groups to perform complex tasks fairly easily, common knowledge can make new ideas difficult to come by and novel experiences difficult to process (Justesen 2004, 89). People tend to cling to what is familiar. As a result, strong ties can create the informational equivalent of

vendor lock-in. Ideally, people and other knowledge gathering systems need *both* sorts of ties: weak ties to provide novelty and encourage innovation, and strong ties to enable complex tasks.

Describing ourselves as “knowledge networks” may seem odd, but is a critical distinction: what and how we know are generated by complex biological and intellectual networks. Textual knowledge is a dense network of sentences, allusions, metaphors, argumentation, disciplinary knowledge, citations to other works, and the knowledge, desires, motives of other human beings, and cultures. While knowledge creation and dissemination has always involved reworking the ideas of others, computer technology’s embedding of cut and paste into the heart of its operation foregrounds the interconnectedness of knowledge (Ulmer 1994). More and more often, we are aware of and encourage involvement in networks of all kinds, both the literal social networks forged at home, work, and play, virtual social networks made possible by the Internet, and those which combine the two, like email lists maintained by academic organizations, or for that matter Myspace and Facebook—which are now so popular they have become uncool (Vara 2006). Catalogs, more than anything, should be about the extension of knowledge networks, and we believe the best way is facilitating the creation of weak ties.

This creation need not be carefully planned and orchestrated. Granovetter points out that because weak ties can be made quickly, we can afford to have many of them without expecting rewards on the short or long term. And different types of ties are more amenable to individuals for different reasons. Full text databases provide an excellent example: some provide better access to summaries and analysis, some to bibliographic information. For certain users, such as undergraduate students, citation data will be less valuable than an explanatory preface or a saved trip to the stacks (Serotkin 2005). For all kinds of cataloging, is difficult to know beforehand what particular connections, or even which kinds of connections, will be the most rewarding on

the short and long term. Hence we believe catalog designers should think carefully about limiting the ways their interfaces function in the name of efficiency or effectiveness, and we suggest emulating the Web 2.0 catalogers who work hard to allow interaction with and integration of networks of all kinds. We'll return to this critical point later in this essay.

Cataloging as communication

All five definitions of cataloging we identified above involve communication, and in fact catalogs themselves are communication media through which libraries and users exchange information. Catalogs deliver not only clues to finding texts one is looking for, but information about cataloging standards and relevant subject headings. Like a language, the catalog itself is a technological infrastructure whose design shapes what can and cannot be represented, not a transparent entity which passes information to users without value judgments. Sanford Berman and Clay Shirky, among others, have pointed out this ideological work in the Dewey Decimal System, which, for example, allots many times the classification space for texts about Christianity as it does for other religions (Shirky 2005). In a more practical vein, Nicholson Baker argued that conversions of card catalogs to OPACs often excluded highly localized and potentially very valuable metadata librarians added to catalog cards, often by hand (Baker 1994).

While it's impossible to create any communication system which functions as in Michael Reddy's "conduit metaphor," where meanings are transmitted as if they were objects unaffected by language, catalogers can make their activities public, so that users can better understand the decisions they make while entering metadata, designing catalog interfaces, and developing standards. In many ways, the publication of library catalog standards such as the Anglo-American Cataloging Rules (AACR2) provides far more openness than the online catalog-like services we are considering in this essay. For example, while Google provides some explanation of its "PageRank" algorithm for ranking web pages, the specifics are kept under wraps (Google

2007). OPAC maintainers should strive to keep this openness, and extend it by integrating more accessible versions of cataloging standards into catalogs of all kinds.

Most library catalogs broadcast information largely in one direction, from the library to the user. Users seldom use OPACs to communicate with libraries and *never* for reaching one another—at least not directly. Campbell and Fast argue that traditional catalogs do act as an important avenue of *indirect* communication between readers and writers by allowing scholars to reliably identify specific works which they may then read and comment upon, with the knowledge that other readers and writers, will do the same with identical texts. Further, they argue that catalogs' function as communication media also allows for fostering communities of practice (Campbell and Fast 2004, 34). Using an OPAC, a scholar researching Coleridge's contribution to Romantic theories of the imagination can easily locate and retrieve a copy of John Livingstone Lowes' classic 1927 study, *The Road to Xanadu*, as well as Coleridge's own *Biographia Literaria*, and enter into dialog with both Lowes and Coleridge—and many other scholars through other books and the citations contained within them. Access to many texts allows scholars to emulate or contest evidentiary and citational protocols found in them. In this way the catalog fosters, indirectly, a community of scholarly practice among readers and writers.

Though the catalog contributes indirectly to this process, its contribution is powerful: remove the ties generated by the catalog, and many research communities would vanish. Certainly, findability remains crucial to research today, though communities of practice are not necessarily forming around or through “proper” library catalogs. At least now, researchers are creating this infrastructure on their own, demonstrating both their desire for two-way inter-research communication, and the ability of catalog-like services to organize it. Publication in many academic scientific communities has changed significantly thanks to sites such as the pre-print server arXiv.org, where articles are posted online in advance of their official publication in

print—allowing the information to be used sooner rather than later (Johnson 2004, 5). Without a doubt, LibraryThing provides a second useful example of a communication-enhanced cataloging service, given its integration of forums and facilitation of contact. Furthermore, scholarly databases such as the Association for Computing Machinery (ACM) Digital Library include authors' email addresses as live links in their records, allowing very rapid and direct communication. For us, these and other examples demonstrate network building through facilitation of weak ties, underscoring our claim that traditional catalog designs must recognize their value as well, and seek to alter their functionality in ways which allow users to create a wide variety of weak ties through the catalog.

Play and parody in cataloging

As cataloging becomes ubiquitous, catalogs need to allow more uses, more contexts of use, and more interoperability with other catalogs and catalog-like systems—in other words, more opportunities for catalogs to be integrated into social, communicative, and educational networks. We have already pointed out some features of catalog-like web sites which we feel OPACs should consider, and we would like to suggest some more. But first we'd like to note two broad movements which demand special attention.

Play and parody should enter the catalog, for they allow for the creation of unexpected ties to known networks, as well as ties to new networks. The boundary space between the known and unknown is where new knowledge is most easily formed. Play, as Jacques Derrida noted, is the result of the inability of any structure to be complete (1978, 289). This is not simply to say all catalogs will be incomplete because there will always be new books but more fundamentally because the catalog is structured around incompleteness and absence. Cataloging assuages an absence, a desire for getting at the knowledge contained in a library's collection and creating new knowledge from it. Catalogs still act as permeable boundaries between people and the

knowledge, “real” and the “potential” knowledge contained in the collection, mediating the indeterminacy between what is known (a work’s title, author, or subject) and the desire for the unknown (the work’s content and, more importantly, its potential use). For example, a reader might know a library has Keith Waldrop’s translation of *Les Fleurs du Mal*, and might even know a little about its content. That same reader may have read reviews or other translations of *Les Fleurs du Mal*. But until she gets and reads Waldrop’s translation, she can’t know what the book knows or how that knowledge will affect her. The catalog was designed to respond to a desire that will never be fully satisfied because it is insatiable, a form of play.

Of course, catalogs and catalogers have always mediated, even wrestled with, the knowledge boundary. Consider how difficult it has been to establish what seems to be known: an author’s name, the correct title, the edition, etc. How much of the unknown to include in the catalog has long been a point of contention. The space constraints of catalog cards (and to a significant degree of Machine-Readable Cataloging (MARC) records after cards), meant that including significantly more information was difficult to do. Not that librarians didn’t try, as Baker’s argument about handwritten notes on catalog cards reminds us (1994, 76).

But traditional cataloging often functions to reduce play. For example, subject access in a library catalog is determined firstly by committees composed of catalogers who approve the subject headings, and secondly by individual catalogers who employ those same headings. The recently created subject heading, “Camp (Style),” will allow a large group of materials formerly dispersed to be collected together. But camp, as a rhetorical and personal style, has been around for a long time, and the thousands of materials published which could usefully be accumulated under this heading will not be gathered there soon. If catalogs, while retaining their own subject headings, opened their discursive spaces, allowing users to classify works as “camp,” and encouraged patrons to share these subject headings with others, a great many items might be

quickly identified as “camp.” Some of these would be items a professional cataloger might not designate as “camp”—for instance, a catalog user might designate Susan Sontag’s “Notes on Camp” as itself a form of camp rather than a description of it. Though seldom admitted into traditional cataloging, these kinds of disagreements are inevitable, interesting, and often fruitful—exactly what we have in mind when we call for the admission of play into the catalog.

Some might argue that changing the boundaries of OPACs in this way would amount to a parody of cataloging, or that it would fail as cataloging. Yes and no. Mikhail Bakhtin points out in *The Dialogic Imagination* that parody is possible because formal discourses, like that of the catalog record, are not complete, not inevitable (1981, 59). Within the limits of their genre (epic poetry, a junior high civics class, or cataloging), such discourses don’t say everything. Parody compels a discourse to say things it ordinarily wouldn’t—perhaps even certain truths. Bakhtin uses the example of Rabelais’s parodies of religion, which say what the church would not say about itself. In our time, “South Park” is a Rabelaisian parody of middle class America, delivered via the forms of normal childhood. (And in 1994 Baker was, in essence, calling the OPAC a parody of the card catalog.) As Bakhtin says, “Parodic travestying literature introduces the permanent corrective of laughter, of a critique on the one-sided seriousness of the lofty direct word, the corrective of a reality that is always richer, more fundamental, and, most importantly, to contradictory and heteroglot to be fit into a high and straightforward genre” (1981, 55). While it might be difficult to accept the notion that parody should be integrated into the texts it refigures, or at the least presented as complimentary, we think its power as critique offers a way for catalogs to simultaneously create the two-way communications we’ve called for above, while facilitating both a very formative strong tie to the original being parodied and valuable weak ties to texts and discourses which would not be connected absent the critique of the parody. Parody can also be an example of allowing play into a system of discourse, another way of telling the

truth, of keeping a discourse honest.

Cataloging in Web 2.0

Our call for catalogs to reinvent themselves by becoming better networks, by integrating communication, and by allowing play and parody is motivated in large part by LibraryThing, del.icio.us, and similar sites which Tim O’ Reilly first labeled “Web 2.0” (2005). Fairly new services such as the social bookmarking manager del.icio.us, or the photo-sharing site Flickr, are joined by comparatively venerable sites like Google and Amazon.com in performing one or more of the five types of cataloging we identify at the start of this essay. Where first-generation web sites used a variety of methods to provide information to users—essentially an update of newspapers or broadcasting—Web 2.0 sites integrate a high degree of user input and interaction. Arguably, that input, whether kept private or made public, whether individual or aggregated, is the core of Web 2.0, and essential to the cataloging functionality of the sites we discuss here. While not all Web 2.0 sites engage in cataloging, many do, and we want to describe why some are so effective, as we continue to draw upon them to suggest methods for rethinking cataloging.

LibraryThing and Amazon.com both allow and even encourage forms of play and parody in several ways. Both allow users to add metadata, usually in the form of keyword-like “tags,” which expand upon the subject headings provided when a record is imported into the system. These tags are searchable across personal collections and across the holdings of both systems. Among other things, Amazon.com allows readers to contribute reviews, comment on other reviews, and create lists of books that are keyed to records. Because it is not standardized this user-contributed content is inherently transdisciplinary, often linking texts in surprising ways. For example, Amazon.com readers’ lists are deeply personal and often irreverent—and very useful because they can connect texts in ways catalogs do not. These lists cover subject matter which might be considered esoteric (for example, information visualization in English studies,

transsexuals in film, Australian rules football) and use a tremendous variety of approaches for evaluation, collecting items not covered by traditional searches. These resources offer additional ways of envisioning texts cataloged by Amazon—without interfering with the site’s own classification systems.

A second example shows the power of parody. LibraryThing’s author and title “clouds” reveal that J. K. Rowling and J. R. R. Tolkien are the most heavily collected authors on the site. Rowling and Tolkien are followed by a host of others, many the authors of genre fiction (mysteries, science fiction, young adult) and, with the notable exception of Jane Austen, few traditionally canonical authors. In other words, the LibraryThing catalog dramatically demonstrates a truth librarians have long known: readers prefer popular fiction to the canon. However, when libraries respond to their users’ wishes and make more space for popular fiction, they are often hammered by traditionalists who perceive the library as an institution meant for the betterment of the public, as recently occurred with the Fairfax (Virginia) County Library’s weeding program (Hollingsworth 2007). In other words, a catalog designed along the lines of LibraryThing could successfully parody the implied cultural elitist view of libraries and open up the age-old debate about libraries’ purposes and clientele. A more participatory catalog would situate this debate between readers themselves, via the catalog (in users groups, for example).

What sites like LibraryThing catalog, and for whom, varies widely, as do the methodologies employed, consistent with our call for a diversity of approaches in the name of facilitating weak ties. Del.icio.us focuses on web bookmarks; anything with a uniform resource locator (URL). Flickr allows uploading, organizing, and sharing of photographs and other computer-generated images. The CiteULike citation manager supports sorting, input, and export of citations for academic writing (and like many Web 2.0 sites, is targeted at a very specific audience: researchers using common databases like JSTOR and bibliographic tools such as

BibTeX). While these and other sites differ in the content of cataloging, they all answer the question “for whom” in two identical ways. The content which users identify, annotate, share, and return to is first intended for the end user doing the work. Bookmark managers like del.icio.us solve a terrible but common problem: how people using multiple computers and web browsers can establish a single source for bookmarks. But there is a critical second audience: anyone using a site which allows users to make their input public can compare her content to individuals who have saved similar items or used similar annotations, as well as to the aggregated display of all site users. In this way, these web services employ what Mike Robinson would call a “double level language.”

Double level languages

Writing about what is now often called “computer supported collaborative work,” Robinson (1991) proposed that applications with a double level language possess a formal language and what he calls a “cultural” language. The formal language structures and organizes the work, providing clarity and the ability to share information at the cost of being rigid. In the case of traditional catalogs, that would be the AACR2 standards; for LibraryThing, del.icio.us, and similar sites, that would be the interface elements which allow users to upload content, add titles and other metadata, and control how it is shared—though we should hasten to note the sites we’ve seen have tried to ensure their rigidity enables, rather than constrains, end users. The cultural language allows work to be commented upon and extended in a manner that does not necessarily sacrifice the integrity of the formal language. On the web, these cultural languages take numerous forms: on- and off-site forums and commentary, patterns in user-contributed information, visualizations of aggregated data, inter-site mashups and combinations, and a general reflection of cultural trends into cataloged content (often called “memes”).

Double language systems draw from the strength of each language. The cultural language

provides the ability to innovate and connect to the “unknown” at the cost of increasing ambiguity or, if you like, semantic entropy. Yet it is this tension between the orderly and chaotic that allows new knowledge to enter into established networks (Robinson 1991, 40, 42–43). Anders Lundquist illustrated this process nicely in a case study of Cisco’s use of newsgroups for product support. Newsgroup users created a cultural commentary on Cisco’s formal work processes and router products, fueling product innovation and also building expertise in its community of users, some of whom even became official Cisco support personnel (Lundquist 2004, 98–99, 102). Most importantly, double-level languages bring together many of the elements we’ve argued for in this essay: they boost the creation of weak ties, encourage two-way communication, and allow play and parody to be integrated without sacrificing the original. In this vein, Amazon.com differentiates between user-submitted and “editorial” reviews; similarly, a catalog site could differentiate between metadata supplied by the Library of Congress and that shared by end users. Furthermore, many Web 2.0 services use rigid formal languages such as Really Simple Syndication (RSS) for information exchange with other web sites, but have few restrictions on the ways these formal languages can be used. This flexibility allows the idea of “double leveling” to become recursive: a list of bookmarks from del.icio.us can be the formal language upon which cultural languages from LibraryThing can build, and so on. Part of this is a matter of attitude: how do site proprietors react when they find their classification systems being used in unintended ways? Those who respond by incorporating users’ innovations, as Lundquist documented with Cisco, are far more likely to develop the energetic, devoted user base capable of such surprising work.

Tagging

While this interaction sounds quite complicated, it’s actually very simple, thanks to the simple but powerful metadata system called “tagging.” Sites which use tags for classification

allow users to create multiple tags, which are a kind of keyword, for contributed content. For example, one can use LibraryThing to create an entry for *Moby-Dick* by tagging it in this manner: “melville, literature, long, novel, whale, tech, toread.” These tags function like subject headings (in that disparate materials may be linked together through them) and can be shared with other users of the service, who in turn may or may not use them for their own purposes. Many sites allow users to copy items from each other including tags and other metadata, in much the same way the Online Computer Library Center (OCLC) facilitates record-sharing. Furthermore, tags don’t have to relate to content fixedly—the tag “toread” is almost certainly temporary and indeed specific to a single user (Shirky 2005).

But tags can provide a quick and easy way to traversing multiple sets of metadata. On del.icio.us, for example, users can quickly move between the following views:

a single item the user dilger has tagged “cataloging”

all items dilger has tagged “cataloging”

all items any user has tagged “cataloging”

a single item tagged “cataloging” by the user wat100

all items tagged by wat100

This rapid movement facilitates the creation of multiple weak ties: to other users with similar tags, to other texts which one might find useful, or to tag lists which one can return to and monitor for new, interesting items. Notably, allowing a catalog to integrate tagging into a catalog would foster connections between items and searches—imagine adding “all items tagged frbr in my OPAC” to the list above. Allowing users to add tags to a catalog, whether integrated in the catalog itself or as a double-level language via external connections to sites such as LibraryThing, would make cataloging a more dynamic and open ended process, a discursive space in which users might create connections or explore connections already made by

professional catalogers. It might even be that tags created by users would be adopted by the professional cataloging community. The catalog could then become a discursive space that allows for the creation of tagging as a form of legitimate peripheral participation in the community of practice of cataloging. Along these lines, del.icio.us automatically suggests tags other users have adopted for similar links. The user can choose to adopt the tags or not. By offering the suggested tags, del.icio.us increases the sociality of the network and fosters the creation of a user driven descriptive language, thus allowing for standardization and idiosyncrasy.

Legitimate peripheral participation is defined by Jean Lave and Etienne Wenger as “the process by which newcomers become part of a community of practice” (Lave and Wenger 1991). Libraries function as communities of practice: newcomers hired into a library learn the “way things are done here,” but because they are on the periphery of a library’s established practices, they are more likely to do things differently, to bring in new knowledge, to innovate. Yet even though they are peripheral, the library is more apt to adopt the innovations of new librarians as opposed to others, because it views them as legitimate (because they have an MLS). Similarly, users who added tags to a catalog would be participating in a peripheral but legitimate way in the work of the catalog, a large part of which is the identification of information and of the relationships between works. Moreover, because users would not be bound by the limits of a particular collection, their tags could lead out of the collection and into other knowledge domains: the holdings of other libraries, journal literature, websites, weblogs, and out of the typographical domain entirely to images, and to audio and video content. This richness of content is visible in the diversity of Web 2.0 sites, some of which privilege the crowd, such as those listed here, but others like Ma.gnolia and Citizendium, which seek to maintain the discreteness of expertise by user-edited groups and peer review structures (Sanger 2006).

Tagging has its critics—librarians and information architects alike have pointed out some of its limitations (Rosenfeld 2005). For example, given that few, if any sites which use tags check users' spelling, using "cataloging" for one or more texts will cause those items to disappear from lists made from properly spelled tags. Darlene Fichter has argued that variant forms such as "bunny, bunnies, and rabbit, Easter, or a proper name" can similarly reduce the usefulness of tagging (Fichter 2006, 43). But these problems have been confronted by traditional catalogers for years—and few would propose we abandoning subject headings because of variations in spellings of "al-Qaida." While tagging and its much ballyhooed companion "folksonomy" (classification by amateurs, as opposed to "taxonomy" created by experts) aren't going to automatically revolutionize the art and craft of cataloging, they remain valuable tools for reaching the larger goals we've set out as cataloging moves towards ubiquity.

Interoperability

There are other ways catalogs could foster weak ties. We are used to thinking of the catalog as one thing and items it catalogs as another, but in a digital environment those distinctions are often blurred. Amazon and Google Books provide limited access to the content of many books in their collections, thus approximating Jorge Luis Borges's vision of a map coincident with the geography it describes (Borges 1972). The limited access Amazon and Google Books provide is through keyword searches of scanned materials and limited browsing of a few pages of any one book. Users can search all pages of all scanned materials in the Google archive, whereas within Amazon this ability is more limited (though books are linked in other ways). These keyword searches make Google Books and Amazon.com powerful research tools. Seeing actual pages from the works, however, leverages this power even further, especially when these pages include a book's bibliography and/or index. Given that information, users can query Open WorldCat from the Google Books record to locate the item in the nearest library, or use a

service like Amazon to purchase the book. Amazon supplements keyword searching with lists of “Statistically Improbable Phrases” (SIPs). According to Amazon, these phrases repeat an “improbable” number of times in a book, as determined by an algorithm which compares all books which have been scanned. (Amazon.com 2006, #3) SIPs are linked to other books which use them, creating a quantitatively driven form of “subject access” that works rather well, though not like traditional subject headings. Instead of directing people toward similar items, SIPs can establish or reveal connections between books which readers might not have expected.

Imagine, the power of these resources if they were made interoperable by the simple method the del.icio.us employs for exchanging information. The web address “<http://del.icio.us/dilger/cataloging>” provides a reverse chronological list of all items the del.icio.us user dilger has tagged “cataloging.” Similar URLs are available for all of the instances we note above, and provide easy access to dynamic metadata sets which are actively maintained. Additionally, one can see recent items *anyone* has tagged “cataloging” with “<http://del.icio.us/tag/cataloging/>”. Reloading that page every few minutes shows the size of the community of people interested in cataloging—all of whom could become nodes in one’s knowledge network. Notably, del.icio.us offers other ways to get lists of tags, such as RSS feeds or JavaScript Object Notation (JSON) objects. Programmers are using these streams of cataloging data to create new ways of organizing the flow of information on the web, drawing upon the double level language capability of many sites to create sophisticated ways of visualizing metadata. We hope for more integration of information from local OPACs into those streams—and for more flow of information to and from the unofficial catalogs of Web 2.0 and the OPACs on whose door they are, ever so quietly, knocking. We hope to see more discussions about all five of the cataloging actives we have discussed here, with participation from academic librarians, programmers working for Silicon Valley startups, and the general public. With the

ubiquitous cataloging we hope to see—a matrix of activities performed with a hybrid of traditional OPACs and other online services—a library patron could learn about book in the library catalog, search its contents via Google Books or Amazon, check a few reviews and make a quick writeup of the text in LibraryThing, and decide whether to purchase the book or retrieve a copy off a nearby shelf.

Bibliography

- Amazon.com. 2006. Amazon.com statistically improbable phrases.
<http://www.amazon.com/gp/search-inside/sipshelp.html/002-2471746-4597601> (accessed October 16, 2006).
- Baker, Nicholson. 1994. Discards: Card catalogs destroyed as online systems grow. *New Yorker*, April 4, 1994, 64–70.
- Bakhtin, Mikhail M. 1981. *The dialogic imagination*, ed. Michael Holquist. Austin: University of Texas Press.
- Bogost, Ian. 2006. *Unit operations: An approach to videogame criticism*. Cambridge: MIT Press.
- Borges, Jorge Luis. 1972. Of exactitude in science. In *A universal history of infamy*. Trans. Norman Thomas di Giovanni. New York: Dutton.
- Byrum, John D. Jr. Recommendations for urgently needed improvement of OPAC and the role of the National Bibliographic Agency in achieving it. World Library and Information Congress: 71th IFLA General Conference and Council, August 14–18, 2005, Oslo, Norway. <http://www.ifla.org/IV/ifla71/papers/124e-Byrum.pdf> (accessed February 12, 2007).
- Campbell, Grant D., and Karl V. Fast. 2004. Panizzi, Lubetzky, and Google: How the modern web environment is reinventing the theory of cataloguing. *Canadian Journal of Information and Library Science* 28 (3): 25–38.
- Derrida, Jacques. 1978. Structure, sign, and play in the discourse of the human sciences. In *Writing and Difference*. Chicago: University of Chicago Press.
- Fichter, Darlene. 2006. Intranet applications for tagging and folksonomies. *Online* 30 (3): 43–45.
- Google. 2007. Corporate information: Technology overview.
<http://www.google.com/corporate/tech.html> (accessed January 17, 2007).
- Granovetter, Mark S. 1973. The strength of weak ties. *American Journal of Sociology* 78 (6): 1360–80.
- . 1983. The strength of weak ties: A network theory revisited. *Sociological Theory* 1: 201–33.
- Greenfield, Adam. 2006. *Everyware: The dawning age of ubiquitous computing*. Berkeley: Peachpit Press.
- Hollingsworth, Helen C. 2006. “Fairfax, don’t take the culling too far; the public library is one

- institution that shouldn't be dumbed down." *Washington Post*, January 21, 2007.
- Johnson, Richard K. 2004. The future of scholarly communication in the humanities: Adaptation or transformation? http://www.arl.org/sparc/pubs/docs/SPARC_CELJ.pdf (accessed February 12, 2007).
- Justesen, Susanne. 2004. Innoversity in communities of practice. In *Knowledge networks: Innovation through communities of practice*, ed. Paul Hildreth and Chris Kimble. London: Idea Group Publishing.
- Lave, Jean, and Etienne Wenger. 1991. *Situated learning: Legitimate peripheral participation*. Learning in Doing: Social, Cognitive, and Computational Perspectives. New York: Cambridge University Press.
- Lossau, Norbert. Search engine technology and digital libraries libraries need to discover the academic internet. *D-Lib Magazine* 10 (6).
<http://www.dlib.org/dlib/june04/lossau/06lossau.html> (accessed February 12, 2007).
- Lundkvist, Anders. 2004. User networks as sources of innovation. In *Knowledge networks: Innovation through communities of practice*, edited by P. Hildreth and C. Kimble. London: Idea Group Publishing.
- Miller, Paul. 2005. Web 2.0: Building the New Library. *Ariadne* 45.
<http://www.ariadne.ac.uk/issue45/miller/> (accessed February 12, 2007).
- Morville, Peter. 2005. *Ambient findability*. Cambridge: O'Reilly.
- O'Reilly, Tim. 2005. What is Web 2.0: Design patterns and business models for the next generation of software. <http://www.oreilly.com/go/web2> (accessed February 12, 2007).
- Sanger, Larry. 2006. Toward a new compendium of knowledge. <http://citizendium.org/essay.html> (accessed February 12, 2007).
- Serotkin, Patricia B. 2005. If we build it, will they come? Electronic journals acceptance and usage patterns. *Libraries and the Academy* 5(4): 497–512.
- Shirky, Clay. 2005. Ontology is Overrated: Categories, Links, and Tags.
http://www.shirky.com/writings/ontology_overnated.html (accessed November 1, 2006).
- Reddy, Michael J. 1993. The conduit metaphor: A case of frame conflict in our language about language. In *Metaphor and Thought, Second Edition*. Cambridge: Cambridge University Press.
- Robinson, Mike. 1991. "Double level languages and co-operative working." *AI and Society* 5 (1): 34-60.

- Rosenfeld, Louis. 2005. "Folksonomies? How about metadata ecologies?"
http://louisrosenfeld.com/home/bloug_archive/000330.html (accessed February 12, 2007).
- Toobin, Jeffrey. 2007. Google's moon shot. *New Yorker* 82 (48): 30–35.
- Ulmer, Gregory L. 1994. *Heuretics: The Logic of Invention*. Baltimore: Johns Hopkins University Press.
- Vara, Vauhini. "MySpace, ByeSpace? Some Users Renounce Social Sites as Too Big." *Wall Street Journal*. October 26, 2006.