



## Is It Worth It? Management Issues Related to Database Quality

Janet Swan Hill MA

To cite this article: Janet Swan Hill MA (2008) Is It Worth It? Management Issues Related to Database Quality, *Cataloging & Classification Quarterly*, 46:1, 5-26, DOI: [10.1080/01639370802182885](https://doi.org/10.1080/01639370802182885)

To link to this article: <http://dx.doi.org/10.1080/01639370802182885>



Published online: 12 Oct 2009.



Submit your article to this journal [↗](#)



Article views: 281



View related articles [↗](#)



Citing articles: 3 View citing articles [↗](#)

# Is It Worth It? Management Issues Related to Database Quality

Janet Swan Hill

**ABSTRACT.** Management issues related to the quality of a library's catalog and its source databases reflect the continual evolution of both. As catalogs seem about to mutate rather than to continue to evolve gradually, a review of how management thought about database quality was developed may assist libraries to recognize which principles and circumstances remain valid and true, to assess which are no longer applicable, and to decide what actions to take. A persistent shortcoming in the decision-making process that needs to be addressed is the lack of serious research into user needs and benefits, and the actual impact on users of database quality decisions.

**KEYWORDS.** Database quality, database management, quality control, catalog management, catalog evolution, discovery tools evolution

## *A HISTORICAL CONTINUUM*

Whatever the area of endeavor, decisions about how to manage it are made to address circumstances that are current at the time of the decision. As the area evolves, issues surrounding it, capabilities to deal with it, and even some of the philosophies underlying it may undergo

---

Janet Swan Hill, MA, is Professor and Associate Director for Technical Services, University of Colorado Libraries, Boulder, CB184, Boulder, CO 80309 (E-mail: Janet.Hill@colorado.edu).

Cataloging & Classification Quarterly, Vol. 46(1) 2008  
Available online at <http://ccq.haworthpress.com>  
© 2008 by The Haworth Press. All rights reserved.  
doi:10.1080/01639370802182885

change, necessitating reassessment of management decisions. As evolution takes place, some issues, philosophies, and decisions may remain unaltered at a basic level while others may lose their relevance, and some entirely new issues may emerge. Over time, the issues to be dealt with, the principles that govern them, and management responses to them build up layer by layer to the point that the rationales that led to their creation or articulation may be obscured. Some actions may be taken so much for granted that they are continued without question far past the time that they have any utility, while others may be airily jettisoned only to give rise to unfortunate consequences because the reasons behind them were not fully understood.

A library's catalog is a complex mechanism that has undergone significant evolution, especially in the last half-century. Because management issues related to its quality and maintenance reflect that evolution, they can be complex and multi-layered. As we look toward a future in which catalogs seem about to mutate instead of continuing to evolve gradually, a chronological review of how the layers of management thought were developed, that relates them to the context in which they were formed, seems especially useful. Understanding the original context and recognizing what circumstances are the same and which have changed can help libraries create intelligent responses to today's challenges.

### ***DEVELOPING THE FOUNDATION-CARD CATALOGS***

Prior to the development of computing, when libraries dealt with paper and manual catalogs, the terms "database" and "database quality" had no meaning, but for ease of expression within this paper, catalogs and other storehouses of data will often be referred to as "databases," whether they were formed of print-on-paper, or had a machine-readable basis.

The foundation of many of today's management decisions relating to database quality were laid down in the context of manual cataloging and card catalogs, in response to their particular characteristics and capabilities. In the era of card catalogs there were very few databases of concern to individual libraries, with the first and most important one being the database that constituted a library's own catalog. When libraries built their own catalogs independently without assistance from outside sources of cataloging, it would have been understood that the maintenance of internal quality was important both to internal recordkeeping

needs such as accurate inventory, and to supporting the traditional objects of the catalog as articulated by Charles Ammi Cutter.<sup>1</sup> Additionally, although the first interpretation of “quality” might have been in terms of accuracy, issues such as usability and completeness would also have been recognized as relevant to overall quality. When it became possible for libraries to supplement their own cataloging by ordering cards from the Library of Congress (LC) or from a commercial vendor and interfiling them with local records, the concept of database quality began to encompass both the quality of individual records, and the quality of the overall catalog structure.

As is true today, management decisions were based on such things as assessing the difference between the ideal and the achievable; deciding whether the cost of a particular activity was justified considering the benefit to users; assessing the negative consequences of not doing something; and determining whether a particular activity was supportable given local resources. Decisions were also influenced by the fact that, for users, there were few to no alternatives to using the tools libraries gave them, so that even if retrieval were awkward or non-intuitive, it was up to the user to deal with it.

Some of the decisions made regarding database quality were philosophical, such as maintaining a catalog that honored Cutter’s objects. Even decisions to include some types of materials in the main catalog (books, journals) while excluding others (e.g., maps, prints, photographs, sound recordings, newspapers) were often based on philosophies about the relative importance of some types of materials as opposed to others. Some decisions were purely practical, such as limiting the number of cards created for individual titles by creating the “rule of three” for subject headings and authors, so as not to exceed the physical capacity of the card catalog.<sup>2</sup> Some decisions were monetary, such as decisions to create brief records for some materials because of inability to fund additional personnel. Some decisions were forced on libraries by circumstances, as when a tight deadline for opening a library forced a decision to provide briefer cataloging than might otherwise have been done. Many policies derived from a combination of these motivations.

Unfortunately, few decisions were supported by research. When it came to assessing user-benefit, libraries often operated from a perspective of “We know what’s good for you, dear,” and relatively little research was done into user desires and capabilities. User-benefit tended to be taken on faith, based on the personal experiences of catalogers, on opinions conveyed by reference librarians, and on a perhaps over-

optimistic assumption that the rules themselves, which had been formulated by august entities such as the American Library Association (ALA) and the Library of Congress (LC), had been established as they were for good and sufficient reason, taking such things as user-benefit into account. In fact, however, reference librarians and technical services librarians were not terribly communicative with one another, each side staying steadfastly out of the other's area of responsibility, and generally regarding such separation as reasonable.

One basic unspoken tenet of database quality in card catalogs may have been derived from the fact that individual libraries had almost total control over the data that went into their catalogs. Much of the cataloging was done locally and originally, and even the possibility of supplementing local cataloging with cards purchased from commercial suppliers or from the Library of Congress did not substantively change the degree of local control. Purchased cards were not available for all that a library might collect, so they did not obviate the need to continue to do cataloging locally. Purchased cards constituted almost the sole source of outside data that would be incorporated into the catalog, and they were often reviewed extensively by professional catalogers, and were revised to match long-standing local practices.

Given these circumstances, many libraries and librarians labored under the delusion that they could achieve something close to perfection in the catalog. Even if it were admitted that actual perfection was not possible, many management decisions were still made with the almost unconscious philosophy that the aim should at least be in that direction. It may have been easier in that time to believe that this was a reasonable goal because errors were generally revealed only by accident as they were stumbled across in the one-user-at-a-time, local-access-only catalog.

The work toward building a perfect catalog began with the creation, selection, or revision of cataloging cards to be included. Individual libraries established what they believed were necessary levels of review to give to original cataloging, and also settled on the degree of review that ought to be given to LC or other externally acquired cards. Libraries also had to decide whether to maintain idiosyncratic local practices and revise external cards accordingly, or whether they ought to revise local practice to increase agreement with purchased cards. Because libraries did not share their own cards with others, and because their cards were not seen outside their own physical premises, decisions to maintain local practice were much more defensible than they might be today.

Once they began to use Library of Congress cataloging, libraries gradually adopted it as the “gold standard” of quality, and regarded LC practice as the best pattern to follow. Since LC did not share documentation regarding its local practices, libraries faced with what looked like a peculiar card from LC would try to divine what principle it represented, even when the peculiarity might have been nothing more than a genuine mistake, or might simply have represented a practice long since obsolete. Libraries also strove for perfection in the order of the catalog, recognizing that a misfiled card could mean a book lost to users, or a book mistakenly purchased twice. They filed cards into the catalog very carefully, reviewing and revising the placement of cards, and requiring less experienced filers to file cards “above the rod”—that is, with the rod that ran through the holes in the bottom of cards already in the drawer still screwed in place. Only after the filing was reviewed by senior staff would the rod be taken out to let the new cards drop into place. Cataloging staff of all levels knew the filing rules intimately, and could sometimes be heard muttering, “People before places before things; nothing before something.” Such an awareness of filing rules might be regarded as a quaint relic of manual systems, but even in today’s online catalog environment where entries are filed automatically by the software, cataloging staff must occasionally make decisions about machine tagging or coding, or about provision of added entries based upon how the machine will file or retrieve the resulting data.

As the proportion of cataloging derived from purchased cards increased, libraries began to consider more seriously whether the levels of review that had been established when most cataloging was original were still appropriate and whether the level of staff devoted to the work could be justified. Having multiple levels of review unquestionably uncovered errors and enabled libraries to increase the quality of their catalog database, and having purchased cards reviewed and revised by professional catalogers could be argued to result in a better overall catalog product, but was the amount of improvement in quality worth the cost?

It was in this atmosphere that libraries perfected the flawed view of cost-benefit analysis that remains with us today. Because it was impossible, or at least very difficult, to assess end-user benefit, benefit to users was frequently left out of the decision-making process altogether. Analyses undertaken tended simply to compare one cost against another, and to define “smaller cost” as a benefit, leading to such analyses being described as cost-benefit analyses. But a cost-benefit analysis asks, “If This approach has a cost of X and the resulting benefit is Y, while the

cost of That approach is Z and the resulting benefit is W, is it better to do This or That?" When, however, the question asked is simply, "Does This approach cost less than That," then what has been done is not a cost-benefit analysis, it is just a cost study.

In the 1960s, another major source of cataloging data began to be published that gave new importance to questions about review and acceptance of external data. The *National Union Catalog* included photographic images of catalog cards produced by the Library of Congress, but also, by including images of cards contributed by libraries across the United States, it provided data for many works which the Library of Congress did not catalog.<sup>3</sup> The cataloging practice represented by these card images in terms of completeness, heading practice, descriptive cataloging standards, etc., was as varied as the contributing institutions, and decisions that had been appropriate when only LC cataloging was being considered might no longer be the best choice.

If, as often happened, a library had a central catalog and also separate departmental catalogs whose cards were not included in the central file, each catalog might have a slightly different set of rules and standards that governed its quality and content, but so long as each catalog was internally consistent, such variances made little difference. If a library had multiple internal contributors to its central catalog, however, another set of issues arose. Must there be mutually agreed upon standards for all contributing departments? If so, who determines them, who polices them, and who bears the responsibility to handle cleanup or reconciliation? Units for which cataloging is not a central activity may have different priorities from one whose primary function is cataloging. Such units may view parts of the process (such as authority work) as extraneous or unaffordable, and may not appreciate the consequences of non-standard practices, especially if it is not they who are responsible for overseeing the overall quality of the file. The resulting tensions among staff and units may inevitably lead to consideration of organizational change. If responsibility for cataloging remains dispersed while oversight and cleanup is devolved upon the central unit, and if the outlying units steadfastly decline to share the same quality standards, the central unit may come to regard the task as Sisyphean, and adjust its own practices accordingly.

These issues were amplified and extended if libraries tried to file in their own catalogs cards representing materials not owned by the library itself, such as cards representing the holdings of the Center for Research Libraries, or cards representing the holdings of other libraries in the immediate area, such as autonomous law, medical, or theological libraries

on a university campus. Non-uniform headings, and uncertainty as to the currency of holdings and location information on the “alien” cards required management decisions as libraries tried to decide whether the “alien” headings would be filed as is with cross references provided, or altered to reflect the host library’s practice; or whether the possibility that a card would remain in the host catalog for an item that had been withdrawn or moved by the contributing library was great enough to justify not including the cards at all.

Concerns such as these were precursors to matters being grappled with in the present day, as libraries are increasingly enabling access to materials not owned or leased by the host library, and as library catalogs provide links to discovery tools that are not at all under the library’s control. When catalogs were largely card-based, or when online catalogs were not connected to the Internet, one set of decisions might have been defensible and enforceable, since the library itself was in control of what the user might find. In today’s context, however, where Internet connectivity enables users to access information from virtually anywhere at no more than a few clicks distance from what might be termed “the local catalog,” the old decisions may at the very least appear to be inconsistent, and adherence to them may no longer be realistic.

### ***THE FIRST EXPERIENCES WITH DATABASES PER SE— SOURCE DATABASES FOR LOCAL CARD CATALOGS***

Moving forward just a few years, libraries entered a period in which they were actually working with databases of machine-readable bibliographic information as a source of catalog cards to file into their still-manual catalogs. At first for most libraries, that database was either a live, dynamically maintained database like OCLC or RLIN<sup>4</sup> or a “static” database like SilverPlatter or BiblioFile<sup>5</sup> that was acquired by subscription, where updated or additional records were only added when the next issue, disc, or compilation was distributed. Library of Congress records were represented on these databases by virtue of their having been purchased by the bibliographic utility or service. A few libraries with their own internal library automation systems bought LC records through direct subscription. No matter how they were acquired, and no matter what other types of records might be contained on the database, LC MARC records, like LC catalog cards retained their hegemony. They were still the gold standard.

With this expansion of sources of data, the concept of database quality also expanded. It still encompassed the quality of individual records (both local records and those in the source database), and the quality (accuracy, extent, and usability) of the overall catalog structure, and it now also considered the degree to which a particular source database met an individual library's particular needs in terms of its collection and user base. Libraries selected source databases partly according to a variety of factors of primarily internal interest, such as cost and compatibility with local technological capabilities, and in answer to such questions as whether the database had good representation of particular needed materials, such as foreign languages, older materials, audiovisual materials, children's materials, etc.

In addition to considering things that might be primarily of concern internally, libraries were also interested in matters that were common to most other libraries, such as the level and completeness of cataloging represented, the accuracy of that cataloging, and the breadth of coverage of the database as a whole. External databases were subjected to widespread scrutiny, and a veritable cottage industry sprang up in which the quality, number, and type of records available in bibliographic utility databases was studied, and management decisions were based upon the results.

In the case of multiple-source databases such as OCLC and RLIN, which contained records contributed by many different libraries, individual libraries developed local lists that established whose records they would accept; whose they would use only with careful review; and whose they would use only as a last resort. Many of these studies were reported in the literature, and their results were eagerly examined and widely used by others. But many remained private and unpublished, as libraries engaged in small, informal studies aimed at their particular interests, and did not think to share their results through publication. Other investigations were undertaken to determine which databases would assure the greatest hit rate of usable records. Major management decisions made as a consequence of these findings included: choice of database or bibliographic utility; extent and manner in which copy cataloging could be employed; and staffing decisions.

Even with the addition of these options for acquiring bibliographic copy, libraries were still in substantial control of their local catalogs. A significant, though shrinking, percentage of the cataloging was done locally and originally, and the catalogs themselves were still accessible only locally. Library of Congress data was still the primary source of non-local cataloging, both because LC was the single greatest contribu-

tor to most source databases, and because if there were a choice of whose cataloging to use, most libraries still preferred LC's cataloging over the cataloging of any other library.

As the proportion of titles to be cataloged for which existing records could be found grew, management decisions concerning database quality became more numerous and more complicated, and the issue of reconciling practice received on copy with practice reflected in original records came starkly to light. When libraries performed most of their own cataloging, cataloging department personnel were predominantly professionally qualified, but as the use of acquired copy increased, the profile of staffing evolved such that support staff were assigned much of the work of handling materials for which bibliographic copy could be found. Libraries gradually came to realize that, from a management point of view, if the majority of cataloging records came from an outside source and were handled by support staff, it was no longer feasible to conduct extensive review of the outside records. Accordingly, it became necessary to make judgments about what elements of the records were most critical to get right, and to concentrate review efforts there.

What mattered most were elements that would hinder or prevent a record being found in the catalog; that would interfere with the gathering function of the catalog; that would hinder or prevent the work's being appropriately identified; and that would hinder or prevent a user from determining that at some level the work being described might be of interest. These elements were generally determined to be the access fields—the entry fields, including title, and the call number. Other elements were recognized to be less important, and things like variations in notes, publication information, level or completeness of cataloging, etc., came to be regarded not so much as mistakes, but as unavoidable consequences of using someone else's cataloging. In some sense, therefore, libraries defined such variations as not being matters of database quality after all.

Meanwhile, the question of authority control began to emerge as a distinct consideration. This aggregation of mechanisms concerned with the control of headings and the imposition of structure on the catalog through cross references had previously been considered an unquestioned and inextricable part of cataloging, upon which the quality of the catalog depended both as to structure, and as to usability. Now libraries began to ask questions such as what headings should be controlled locally, at what point in the workflow the control should be exercised, and by whom. Once libraries began to view authority control as a separable aspect of cataloging, the door was opened to questions such as whether

to rely on some outside library (the one from which the library obtained most of its copy, for instance) to have done acceptable authority control, and thus to release the local library from the obligation to do the work for itself.

For libraries that continued to maintain a card catalog and that had no internal library automation system, the only time that staff saw an actual machine-readable database was when they were looking at the source database.<sup>6</sup> For such libraries, the primary factor when assessing the quality of the external source database was the impact that it had on the ability of the library to maintain a good catalog locally. For libraries that utilized an internal library automation system in which staff could view records in the context of a full database, even though the public catalog might still have been in card form, the impact that poorly controlled data had upon local ability to do work expeditiously and well could be more clearly seen.

### ***THE LOCAL DATABASE AND THE SHARED ONE- LOCAL ONLINE CATALOGS WITH UNITARY SOURCE BIBLIOGRAPHIC DATABASES***

And so, libraries slid into the era of online catalogs, but most still utilized one primary source for externally acquired records. By the early 1980s a number of libraries had developed local or shared databases that functioned in some respects as catalogs. The use of and sophistication of these catalogs expanded rapidly, and as more libraries built their own databases the question of database quality continued to become more involved, especially because the quality of a catalog in online form was easier to assess than had been the case with cards. Instead of seeing isolated bibliographic records one at a time, in a local machine-readable database individual records could more easily be seen in the context of the records around them, and in the context of the catalog as a whole. As the one-user-at-a-time card catalog gave way to the many-simultaneous-users online catalog, not only did the individual eye see more records, but there were more eyes looking, so more errors came to light. Libraries could also now perform systematic searches for persistent misspellings or investigate suspect patterns in practice through targeted "error hunts."<sup>7</sup> As systems became more sophisticated, so did mechanisms for finding and gathering errors for correction through doing "find" operations; creating lists drawn from the database; and through

automatic generation of lists of possible problems through mechanisms such as new heading notification systems.

Libraries that had subscribed to the philosophy that every error must be fixed found that following this philosophy in the face of more errors discovered placed new strains on staffing. Some responded organizationally by creating special database maintenance units or by repurposing existing catalog maintenance units. Others retrenched.

Now that online access was accepted as the future, new issues arose. Given that completeness of the database has a bearing on its overall quality, libraries began to address expanding the database through incorporation of records for materials previously excluded from the main catalog; materials previously unrepresented in any catalog; and materials cataloged only in manual form.

Retrospective conversion, or translating previous manual cataloging into machine-readable form so it could be included in the online catalog database became a major management issue. At first, when the card catalog was still part of everyone's library experience and expectation, and when the portion of a library's collection that was represented by machine-readable records was relatively small, it was possible to believe that retrospective conversion was an unaffordable luxury, or that users' needs could be satisfied with only selective conversion. As the portion of the collection that was represented only manually decreased, however, and as the portion of the user base that had experience with card catalogs gradually dwindled, it began to be seen that manually cataloged materials were not being accessed or used, and the imperative for retrospective conversion grew. The decision was no longer whether to do it, but how to fund it (through reassigning current staff, through special grants); how to do it (with local staff or through outsourcing on site, or through sending cards or copies of cards off-site); how much to do (everything; only some parts of the collection; only materials of a certain age); and in what order.

Other critical decisions addressed the quality of individual records. Should the library aim to create retrospectively converted records at the same level of quality and review imposed during current cataloging? The expense of doing retrospective conversion led many libraries to choose instead a "quick and dirty" approach to conversion, creating partial records, limiting entry fields traced, imposing limited quality control over outsourced work, etc. Even if they later regretted some of their decisions, few libraries had the resources to revisit the work. Just as studies of database quality and hit rate had inspired a spate of publication, retrospective conversion projects began to be described in the liter-

ature, but most papers concentrated on how and how much, rather than addressing user impact.

The experience of developing projects to accomplish retrospective conversion provided libraries with a perspective from which to view other similar mass projects to increase database coverage by adding records for materials previously excluded from the central file, and by creating records for materials previously uncataloged. Such projects involved the same kinds of decisions that had had to be made for retrospective conversion, such as establishing priorities, deciding whether or not to do the work locally, what level of cataloging records to use, and what kinds of quality control to impose.

Within libraries, maintaining isolated catalogs for certain collections or materials came to be seen as less desirable and defensible at the same time that inclusion of records for these previously ostracized materials became easier. Thus, more libraries began to grapple with quality issues that accompanied building and maintaining a single catalog with multiple contributors.

As soon as libraries began to participate in bibliographic networks and to contribute their own cataloging and holdings to a database that many other libraries used, a sense of community and belonging began to develop. Libraries had accepted contractual obligations for participating in the shared environment, but other issues began to be recognized. As the number of libraries that shared the source databases increased, and the proportion of records contributed by member libraries grew, so did awareness of some issues that were partly practical, and partly ethical, and touched on the question of what obligations one library might have to others. What would the consequences be if some contributors to the shared resource decided not to uphold the common quality standards, even though they might have the resources to do so? What would the impact be if some libraries decided not to contribute to the growth of content but instead to set aside works for which copy was not yet available until another library performed original cataloging? What are the ethical principles that a library should take into account before adjusting its local practices with an eye toward sharing as little as possible in the burden of creating the shared database?

Other cooperative arrangements that began to be introduced provided libraries with more opportunities to share the work of creating more complete databases of a higher quality, which required libraries to address similar questions. Programs such as the Program for Cooperative Cataloging (PCC) and its component parts CONSER, NACO, SACO, and BIBCO, as well as OCLC's Enhance<sup>8</sup> all enabled libraries

to improve what was coming to be called “the national database.” For the most part, the benefits of contributing to these programs were intangible, while the effort involved might be considerable. A decision to become a participant in any such program needed to weigh the burden of undergoing extensive training and review and perhaps accepting the responsibility to contribute a minimum amount of work each year, against such benefits as a deeper understanding of cataloging rules and Library of Congress operations; the honor of belonging to a select group; the satisfaction of contributing to the greater good; the value of enhancing the skills and perspective of local personnel; and the possibility of having some influence over the present and future of bibliographic control.

From the beginning, as soon as libraries were able to acquire catalog cards (and later on cataloging copy), for a significant portion of the materials they needed to catalog, a shift began in staffing and work assignments, and continued until only the most difficult and sophisticated work was done by librarians, while the rest was handled by support staff. With libraries utilizing major online source databases and operating online catalogs, the trend became so obvious that some saw it as marking the deprofessionalization of cataloging itself.<sup>9</sup> Other questions began to be asked: If librarians are not actually used to create cataloging data, where will they learn what they need to know to make informed decisions about the catalog—its structure, integrity, capabilities, content, and even its evolution? Where will they learn what they need to know to enable them to make good decisions about choosing an integrated library system? Where will they learn what needs to be done to imported data to make it compatible with the home database as to quality and content? Do some libraries have an ethical obligation to maintain an organization that preserves their institutions as learning places for those who will be responsible for assuring and recognizing database quality and for imagining and designing mechanisms that will make it possible for all libraries to maintain databases at a level of quality that matches user needs?

### ***LOSING CONTROL OF THE DATA— LOCAL ONLINE CATALOGS WITH MULTIPLE SOURCES OF BIBLIOGRAPHIC RECORDS***

The next stage of catalog evolution brought libraries almost to the present day, and many of the situations described in this section are part of current library operations, as challenges to setting an appropriate, de-

sirable, and implementable course for database quality continue to build. Although text in the previous section was phrased as if libraries had only two main sources of data for their catalogs: their own cataloging, and (at first) LC cataloging, or (later) LC cataloging plus bibliographic network member cataloging data, this is a greatly simplified view of what was actually available. Other sources of cataloging data gradually began to be added, from manual sources such as the National Union Catalog, to other aggregations of data acquired by and mounted on bibliographic networks, such as UKMARC records for materials cataloged in the United Kingdom, and Marcive records for United States government documents.<sup>10</sup> Other sources of data included records supplied by library material vendors at the time of purchase, records acquired with major collections of microforms; records acquired through outsourcing cataloging; records downloadable from the Internet, and many more. Each category of records that became available posed its own challenges, as each might have been created under different standards, for different purposes, and with a different level of quality control. With the addition of each category, libraries had to adjust their decisions about copy acceptance and review.

Whereas in the past, database maintenance and quality control might have been handled on a more or less routine level as records were added to the local catalog one by one, now records were likely to be dumped into the databases in large batches. Processes and policies that had been able to cope with a steady though high level of quality control activities could be overwhelmed by new circumstances in which the previous "normal" demand for quality control was augmented by workload of an unpredictable type and amount at unpredictable times. Budgeting staff for database maintenance became almost impossible. Libraries that had previously rejected outsourcing activities such as authority control began to reconsider, and the quality, flexibility, and cost of outsourcing vendors became a management issue. Libraries that had tried to maintain the same level of review and control of their databases as they had over their card catalogs began to ask whether that level was possible any more, and to lower their sights.

Libraries had once had almost total control over the data in their catalogs, but that control was diminishing. Too many records were coming from too many different sources, with records formulated under too many different sets of understandings regarding database quality. Electronic information resources exacerbated the problem. Not only did they represent new types of material, but they were often described and controlled by means of new sets of rules, or "schema," and the resulting

data might even have been provided by entities or persons outside the normal bibliographic establishment.

A special problem was presented by the prospect of system migration. Some of the older generation of integrated library systems (ILS) were being discontinued or no longer supported, or were consolidating with other systems, while new ILS generations offered new capabilities. Systems that existed outside the ILS addressed newfound needs, such as maintaining digital visual archives or institutional repositories. As libraries contemplated moving some or all data to new systems, they needed to determine whether or not the old records were compatible with the new system. The older the system being moved from, and the longer the library had been using it, the greater the likelihood that there would be some record incompatibility. If records or practices were incompatible, libraries had to come to an understanding of what the impact would be on the new system, and decide whether or how the problems could be fixed either before or after migration, and whether the benefit would repay the expense. Libraries that had experienced a rocky system migration in the past were likely to understand the importance of developing policies regarding record content and structure that would not hinder future system migration. Libraries that had not benefited from a similar experience were often treated to horror stories by their colleagues, and advised to learn from others' suffering, and to pay attention to this particular aspect of database quality.

As these and other challenges emerged, it became necessary for libraries to reexamine long-standing policies and expectations governing the relative degree of data perfection that it is advisable and feasible to aim for. Unfortunately, there was and is still a dearth of research to help libraries understand the impact of individual decisions on user-benefit and catalog usability. Libraries are still making many decisions almost intuitively, and are still primarily examining costs.

An unanticipated effect of the shifting profile of staff engaged in building and maintaining databases away from librarians and toward support staff was to widen the gulf between people who have a detailed understanding of the catalog data, the database structure, and the interplay between them, and those who don't. Those whose familiarity is primarily as an end-user often have little awareness of the work that goes into assuring a usable catalog, and may argue either for the addition of mechanisms to increase the quality (especially the extent) of database content, or for the elimination of routines whose impact is not obvious to them. Thus, even communication is a management issue related to database quality. As rules, standards, mechanisms, materials, and ex-

pectations keep changing, continuing education also emerges as a quality control issue.

***THE SLIPPERY SLOPE—  
LOCAL ONLINE INTEGRATED DISCOVERY TOOLS  
AND BEYOND***

The catalog described in the previous section could clearly trace its ancestry back through generations to card catalogs that were created and maintained locally. The construct that libraries are moving toward today, that we still most often call by the name of “catalog,” appears to have skipped several generations, or at least to be moving through generations at a bewildering pace. It has evolved beyond a single database composed of one type of records to a gateway to information, local and worldwide. It incorporates or provides access to a vast variety of information resources, physical and virtual—either surrogates to the full resource, or increasingly the complete resource itself. It accommodates variation in record content and record structure, created by an assortment of entities according to a broad array of standards. This new catalog can perform searches of great sophistication over immense stores of data at unfathomable speed, but even those searches do not always yield a useful result.

In this new discovery tool, the portion of information for which a local library (or any library) may have been responsible is small, and getting smaller. Instead of having to use what libraries and their practices have traditionally offered in the way of finding tools, users now have a choice of how to seek information. They also bring a different suite of skills and expectations to their search, and have different ideas about what should be available to them. The “we know what’s good for you, dear” approach to database creation and quality no longer works.

No longer can a single library or group of libraries exercise meaningful control over all of the data that users may gain access to. Philosophies, principles, and management decisions that originated in the context of catalogs that used a limited number of strict standards must be reexamined. Many remain valid, but many may not, and some are now difficult or even impossible to support. But as we consider change, it is important not to assume that a principle or practice is obsolete just because it is old, and not to assume that just because something costs a lot that it costs too much. In any reassessment of practice, libraries need to understand why past practices were developed, what principles they

supported, and what purposes they served. Now more than ever it may be wise to consult Cutter and Ranganathan, and decide whether the principles that have guided us for so long should continue to be a part of our vision and purpose.

All of the old management issues, such as the amount of review to give to original and copy cataloging, and which types of data are most critical to control, remain. Others are new, or have taken on new urgency. These include such matters as whether to bring alien records into the central database and whether to revise them to fit better into the catalog aspect of the central discovery tool; how much control the individual library can afford to exercise; how much control is feasible from a technological point of view; whether solutions to some problems can be handled through software or whether there are still areas where human intervention is required; and how much control is really necessary, given new searching capabilities and changes in user behavior. So powerful are some of these new searching capabilities that many both inside and outside the field are asking whether, with keyword searching, searching across databases, full-text searching and retrieval, aggregated searching, and massive combined databases such as Google, if "legacy" systems such as authority control, authorized thesauri are still relevant, and whether quality control of the content of records is still necessary. As impressive as these new searching mechanisms are, however, they still have substantial shortcomings,<sup>11</sup> and even eager proponents and developers of the new discovery tools concede that the best retrieval may depend on the availability of more data, not less, and of data that is authoritative and controlled.<sup>12</sup>

Those responsible for library databases must therefore ask whether the prodigious capabilities of the new discovery tools are sufficient to override their substantial shortcomings, and whether the willingness of today's new generation of users to sacrifice such qualities as consistency, specificity, and gathering in favor of speed, convenience, and transparency is sufficient permission for abandonment of past practices and goals. Ultimately, if machine searching capabilities continue to improve, libraries may find themselves needing to ask whether the concept of database quality can still reasonably encompass accuracy of individual records and whether, instead, extent and content of individual records, extent and content of the database as a whole, and the effectiveness and accuracy of mechanisms to expose those records and that database to the World Wide Web have become the real measures of database quality.

In April, 2006, in announcing that in ten days it would cease creating series authority records, the Library of Congress took what many believed to be merely the first of many potential major changes to its cataloging practice.<sup>13</sup> This expectation was bolstered by the content of "The Changing Nature of the Catalog and Its Integration with Other Discovery Tools,"<sup>14</sup> commissioned by LC, which questions the viability of such activities as continuing past subject heading practice. LC is not alone in considering radical change. Other reports that received wide circulation were "Rethinking How We Provide Bibliographic Services for the University of California," which is highly critical of the existing catalog construct, and suggests ways in which its usefulness, responsiveness and usability need to be improved,<sup>15</sup> and "A White Paper on the Future of Cataloging at Indiana University."<sup>16</sup>

As thought-provoking as these reports might have been, it was LC's announced and possible changes that caused the greatest immediate concern and inspired quick response from a wide variety of librarians and professional associations.<sup>17</sup> Further, as the responses made clear, libraries were concerned with more than the series authority decision itself. They were disturbed by the manner in which the decision had been announced (without prior consultation with the field, and without adequate warning to enable libraries to develop strategies to deal with the result), and by the possibility that this decision was only the first of many that might be made by LC about the cataloging services it provides to other libraries. Libraries feared that the Library of Congress was abandoning its position as a leader in development of standards and practice for providing bibliographic control, as the source of bibliographic records produced at the highest standard of quality, and as the primary source of bibliographic records in terms of numbers alone.

LC's public statements,<sup>18</sup> including its initial proposal to convene a summit meeting in 2007<sup>19</sup> on the future of bibliographic control seemed also to indicate that the Library of Congress was looking far beyond making changes in practice related to the current catalog. Instead it appeared to be looking to a future in which the search systems of the past centuries were abandoned; where libraries ceded to commercial entities the guidance of future information access mechanisms; and in which human efforts and human judgment could be supplanted in whole or in part. Libraries' traditional vision of serving all users for all purposes seemed to be implicitly replaced with a vision of serving the needs of

an “average” user with relatively simple demands and relatively low expectations of service.

Across the United States, libraries of all types have predicated their database practices and staffing patterns on the supposition that LC would continue to catalog materials at a certain level in terms of numbers and quality, and in a manner suited to the current version of the library catalog. If LC ceases this work, then database quality issues will become thornier than ever, leading libraries to ask questions such as: Can individual libraries or consortia possibly take up the work that LC ceases to do? How much responsibility do individual libraries have for doing so? Do entities such as Google and Microsoft really understand library user needs? Will commercial entities address the “common denominator” and neglect the needs of those on the up- and downslopes of the bell curve? If the Library of Congress can no longer support libraries and bibliographic control, who will? How much influence can libraries or associations have over LC, or Google and other commercial entities?

Contrary to their stereotype as inflexible and unquestioning guardians of the status quo, librarians have shown over the last half-century, that they are forward-looking, and eager to adopt new technologies that enhance their ability to provide service. So forward-looking are librarians that many will embrace and discuss future possibilities as if they were already implemented. Especially at times of major transition such as face us today, it is important that libraries be guided by both visionary enthusiasts and doubting pragmatists. It is critical that what works today not be abandoned before replacements that are accessible to the majority have been demonstrated to function well for all purposes of the catalog. It is crucial that replacements be created and evaluated based upon real knowledge of what users want, what they need, and what they will be satisfied with. It is also important that libraries as well as users assess whether the new tools really do serve user purposes, or only appear to.

In order to know these things, libraries can no longer neglect serious research into the characteristics and needs of all users. They must come to a better understanding of the impact of current and proposed practices on the quality and capabilities of the new discovery tools, including the catalog. Assessment of user-benefit needs to become as common in library and information science (LIS) research, publication, and discussions as cost. Both LIS faculty and library practitioners must engage in this kind of inquiry, and professional associations and commercial entities that hope to play a part in the future of information discovery must support it through targeted re-

search and funding. Most importantly, perhaps, those engaged in the research effort must not suppose that they know the answers in advance, and must do work that seeks the real answer, and does not seek to support what it is hoped that the answer will be.

*Received:* April, 2007

*Revised:* June, 2007

*Accepted:* September, 2007

## NOTES

1. The objectives of the catalog were first explicitly stated in 1876 by Charles Ammi Cutter in his *Rules for a Dictionary Catalog*, (Washington, D.C., U.S. Government Printing Office. 4th ed. 1904). These rules stated that a catalog should enable a person to (1) find a book of which the author, or title, or subject is known; (2) to show what the library has by a given author, on a given subject, or in a given kind of literature; and (3) to assist in the choice of a book as to its edition (bibliographically), and as to its character (literary or topical). "Cutter's Objects," as they came to be called, were later joined by the "Five Laws of Library Science" as formulated by S. R. Ranganathan in 1931. These laws state that (1) Books are for use; (2) every reader has his or her book; (3) every book has its reader; (4) save the time of the reader; (5) the library is a growing organism. S. R. Ranganathan, *The Five Laws of Library Science*. Madras Library Association (Madras, India) and Edward Goldston (London, UK). Available: <http://dlist.sir.arizona.edu/1220/> (accessed July 31, 2007).

2. The "Rule of three" is the instruction that only three subject headings should be assigned to any one work. If more than three specific subjects are needed to convey subject content, then instead of assigning three specific headings, the cataloger is to assign a general subject heading that encompasses the narrower topics. Thus, a work that discusses lemons, limes, and oranges would have one subject heading for each fruit, while a work that discusses lemons, limes, oranges, and tangerines would receive a single subject heading for citrus.

3. *National union catalog, pre-1956 imprints*. A cumulative author list representing Library of Congress printed cards and titles reported by other American libraries ([London]: Mansell, 1968-80). This 685-volume set was continued by 125 volumes of the *National Union Catalog, 1956-1967*.

4. Founded in 1967 as the Ohio College Library Center, OCLC (<http://www.oclc.org/>) developed a computerized system by which the libraries could share cataloging data. OCLC now operates the most widely used bibliographic utility in the world, with a combined membership of 41,555 libraries in 112 countries and territories around the world. The Research Libraries Information Network (RLIN) is the bibliographic utility run by the Research Libraries Group (<http://www.rlg.org/>). Originally the membership profile of the two organizations was quite different, with RLIN contributors being limited to academic libraries, and materials represented in the two databases

reflected that difference. In 2006, OCLC and RLG agreed to merge (web sites accessed July 31, 2007).

5. Both BiblioFile and SilverPlatter are CD-ROM based. BiblioFile is part of TLC (<http://www.tlcdelivers.com/tlc/tlchistory.asp>). SilverPlatter is part of the Ovid product line (<http://www.ovid.com/site/index.jsp>) (web sites accessed July 31, 2007).

6. Even those who used a machine-readable source database might never have seen more than one record at a time, depending on how records were searched and retrieved.

7. Librarians still engage in error hunts, some assisted by the Web site "Typo of the Day" which proclaims itself "Database Protectors, Worldwide: 160 librarians working together for better data." Available: <http://typooftheday.blogspot.com/> (accessed July 31, 2007).

8. The component parts of PCC are described on its home page. CONSER and BIBCO enable libraries to contribute serial and monographic bibliographic records to the national database; NACO and SACO enable contribution of name and subject authority records. <http://www.loc.gov/catdir/pcc/>. OCLC's "Enhance Program is designed to allow skilled catalogers to improve the quality of the OCLC WorldCat database by upgrading WorldCat records, primarily from less-than-full level to full level." <http://www.nelinet.net/oclc/cataloging/enhance.htm> (web sites accessed July 31, 2007).

9. Ruth Hafter, *Academic Librarians and Cataloging Networks: Visibility, Quality Control, and Professional Status*. Contributions in Librarianship and Information Science, no. 57 (New York: Greenwood, 1986).

10. <http://www.marcive.com/HOMEPAGE/WEB1.HTMMarcive> (accessed July 31, 2007).

11. Martha M. Yee, with a great deal of help from Michael Gorman, "Will The Response Of The Library Profession To The Internet Be Self-Immolation?" Available: <http://slc.bc.ca/response.htm> (accessed July 31, 2007).

12. Nancy J. Fallgren, "Users and Uses of Bibliographic Data Meeting, March 8, 2007 Mountain View, CA, Brief Meeting Summary." Available: [http://www.loc.gov/bibliographic-future/meetings/2007\\_mar08.html](http://www.loc.gov/bibliographic-future/meetings/2007_mar08.html) (accessed July 31, 2007). Summarizes the first public meeting of the Library of Congress Working Group on the Future of Bibliographic Control.

13. The original date of implementation was May 1, 2006. This date was changed to June 1 in response to a public outcry. The updated announcement was posted at: <http://www.loc.gov/catdir/series.html> (accessed Oct. 22, 2006).

14. Karen Calhoun, "The Changing Nature of the Catalog and Its Integration with Other Discovery Tools" Final Report, March 17, 2006. Available: <http://www.loc.gov/catdir/calhoun-report-final.pdf> (accessed July 31, 2007).

15. University of California Libraries Bibliographic Services Task Force, "Rethinking How We Provide Bibliographic Services for the University of California." Final Report, December, 2005. Available: <http://libraries.universityofcalifornia.edu/sopag/BSTF/Final.pdf> (accessed July 31, 2007).

16. Task Group on the Future of Cataloging at Indiana University. "A White Paper on the Future of Cataloging at Indiana University." January 15, 2006. Available: <http://>

[www.iub.edu/~libtserv/pub/Future\\_of\\_Cataloging\\_White\\_Paper.pdf](http://www.iub.edu/~libtserv/pub/Future_of_Cataloging_White_Paper.pdf) (accessed July 31, 2007).

17. An online petition, entitled "Prevent the Library of Congress from Abandoning the Creation of Series Authority Records," launched by Elaine R. Sanchez. Available: <http://www.petitiononline.com/MARC830/petition.html> (accessed July 31, 2007) garnered 3495 signatures. Statements issued opposing LC's action or urging delay, reconsideration, or cooperation included: "Statement concerning Bibliographic/Cataloging Services at the Library of Congress" from the American Library Association Executive Board, available: <http://www.ala.org/ala/alctscontent/alctspubsbucket/bibcontrol/ALABoard.htm> (accessed July 31, 2007); "ALCTS issues statement on the Library of Congress series authority record decision" from the Association for Library Collections and Technical Services, available: <http://www.ala.org/Template.cfm?Section=pressreleases&template=/contentmanagement/contentdisplay.cfm&ContentID=128028> (accessed July 31, 2007); "Resolution on the Library of Congress Management's Decision to Cease the Production of Series Authority Records" from the Library of Congress Professional Guild, available: <http://www.guild2910.org/SARResolution.pdf> (accessed July 31, 2007). Letters or e-mails were also sent from the Special Libraries Association and the American Association of Law Libraries (jointly), from the Art Libraries Society of North America Cataloging Advisory Committee, and the Africana Librarians Council, the texts of which were posted to the cataloging electronic discussion list, AUTOCAT. The Program for Cooperative Cataloging issued a statement that "formally recognizes and supports the right of the Library of Congress (LC) to make cataloging decisions in its own best interest. . . . and is unwilling to take a stand against LC's decision to discontinue authority control," available: <http://www.loc.gov/catdir/pcc/seriesPCC.html> (accessed July 31, 2007), and the Association for Research Libraries endorsed the LC decision. Meanwhile, OCLC posted "OCLC's Response to the Library of Congress Decision." Available: <http://www.oclc.org/news/announcements/announcement191.htm> (accessed July 31, 2007). A paper by Library of Congress reference librarian, Thomas Mann, prepared for the Library of Congress Professional Guild entitled, "What Is Going on at the Library of Congress?" was posted to the Guild's website. Available: <http://www.guild2910.org/AFSCMEWhatIsGoingOn.pdf> (accessed July 31, 2007).

18. Statements such as this: Deanna B. Markum, "The Future of Cataloging," Address to the Ebsco Leadership Seminar Boston, Massachusetts January 16, 2005. Available: <http://www.loc.gov/library/reports/CatalogingSpeech.pdf> (accessed July 31, 2007).

19. The Library of Congress modified its proposal for a summit meeting, and instead, created the Library of Congress Working Group on the Future of Bibliographic Control, whose charge is to "Present findings on how bibliographic control and other descriptive practices can effectively support management of and access to library materials in the evolving information and technology environment; Recommend ways in which the library community can collectively move toward achieving this vision; Advise the Library of Congress on its role and priorities." Through a series of public hearings, receipt of testimony, its own research and deliberations, the Working Group is to prepare a report to be delivered to the Library of Congress before the end of 2007. Information about the Working Group, its membership, agenda, schedule, background documents, and findings is available at: <http://www.loc.gov/bibliographic-future/> (accessed July 31, 2007).