

Looking Back--Looking Forward: Reflections of a Transitional Librarian

Diane I. Hillmann

Introduction

When I look back at my thirty-five years in libraries, it seems obvious that the wiring in my brain was formed by my training and experience as a librarian, more particularly as a cataloger. I had the benefit of starting out with an education in broadcasting, then after several years of making a living in a library--and deciding that there were no jobs available to women in radio that I cared to do--I went to library school. As a consequence of that late change of direction, and the fact that I worked full time and went to school part time, I managed to work as a support staffer in almost every library department: acquisitions, cataloging, circulation, and special collections. Although this seemed at the time a rather circuitous and cumbersome route to a professional career, I learned essential lessons on how things worked in every part of the library, which I would have missed had I taken a more direct path.

My cataloging career began in 1969 in the Syracuse University Audio Archives, describing 78 r.p.m. recordings using a manual typewriter. From there I cataloged materials in many languages (some of which I neither spoke nor read), and many formats, especially serials. In doing these tasks, I graduated from manual to electric typewriters (with card platens!) to OCLC terminals, to personal computers with client/server software. I freely admit that though I have cataloged almost every kind of material available in a library, I was not and could never be a "catalogers' cataloger." I lacked the interest in rules and details that drives the best catalogers, so I became a

manager of catalogers and technical services. As a manager, I was a “recreational cataloger”—willing, and sometimes eager, to take an occasional break from management and get my hands on real library materials.

Another career shift in the mid-1990’s to become Authorities Librarian for Cornell University Library required an additional step back from the traditional cataloger focus on individual records (albeit records designed to fit into an extensive catalog) to a view of the catalog as an aggregated access mechanism enhanced by consistent and predictable points of entry. Because I was also charged with overseeing import and export to the catalog database, database-level record quality, and reliability concerns crept inexorably into my worldview. These issues became critical ones as the library began planning to migrate MARC records—its most important investment—to a new database under the aegis of a new library management system. This was the third time I had participated in the move of Cornell’s MARC database, and, I hoped, the last.

The view I had developed over this time period—of what worked and what did not, which strategies paid off and which did not—gave me an important perspective as I moved into the new world of metadata. There was a great deal worth keeping from the relatively closed world of MARC, as the new, open world of metadata developed and grew. I felt I was in the right place and right time to participate, with the benefit of experience in the library world to keep me honest and grounded in reality.

Embedded Librarians

Since 2000, I have become engaged in a far different culture than the one in which I began, working for the National Science Digital Library (NSDL) as an “embedded librarian” in an environment heavily populated by computer scientists

whose view of the emerging digital culture is far different from mine. Different languages are spoken here, and different assumptions underlie the conversations taking place. Computer scientists, even those who see themselves as “information scientists,” do not necessarily believe, as I do, that the experience and knowledge of librarians translates well into the digital library world. This is at once a challenge and a frustration, and an opportunity for growth for those of us who find ourselves living amongst those who think that “information science” began with the computer.

Many of the differences in culture and approach that exist are perhaps best expressed as a difference in goals. Librarians are committed to maintaining access to legacy materials and services even as new resources and services are being developed. Their emphasis has always been towards direct user services, visualizing the human being at the end of the equation. Librarians take their position as advocates for their users very seriously, and if they participate in research and development, it is generally with a concrete notion of what should result from that effort. In contrast, computer scientists value research primarily and have little experience in providing services to users. In their world, new ideas result in proofs of concept, papers, and seminars—not production-level services. There are no “end users” in their world, nor legacy systems and legacy data to drag along like an overstuffed backpack. In the world of computer science research, there is little value placed on building concrete services, and those that actually get far enough to be used are often abandoned, or supplanted by even newer ideas that seem worth pursuing. In libraries, ever tighter budgets reinforce librarians’ traditional disinclination to experimentation, as new services tend to attract users quickly, and once in place both are hard to dislodge.

As Thomas Bruce and I point out in our chapter “The Continuum of Metadata Quality: Defining, Expressing, Exploiting,”

There are subtle difficulties rooted in the deeply embedded and divergent expectations that library and computing communities have about audiences, the permanence or persistence of metadata design decisions, and the stability of technology. The cataloging of printed materials is generally done with the expectation that metadata creation is a one-time proposition. Library catalog records are seldom revisited, and new views of metadata are seldom created unless there is great economic or political incentive to do so.

By contrast, computer technologists come from a world in which techniques are continually changing and often improving. This encourages a more experimental, iterative approach to metadata extraction and other machine-processing efforts.¹

Thus we see two worlds: a library world that is inherently conservative and cautious, moving into digital libraries carefully and deliberately, emphasizing the creation of user services based on tried and true methods; and the world of computer scientists, which is more interested in experimenting with new technologies and methods, their primary audience being a small community of like-minded academic researchers. In this second world, the notion of a production-level service maintained over time is a foreign concept. This group tends to see the rapid change in technology as an opportunity without a downside; while librarians also see the opportunities but yearn for the stability to explore them fully and exploit them when carefully justified by experience.

Seeking middle ground in this cultural divide are those who come from library culture but are impatient with its inability to quickly embrace the new and exciting. They disparage legacy library practices, excoriate catalogers for their devotion to detail and wariness of new metadata formats, and far too often promote a variety of new

¹ Thomas R. Bruce and Diane I. Hillmann, “The Continuum of Metadata Quality: Defining, Expressing, Exploiting,” in *Metadata in Practice* ed. Diane I. Hillmann and Elaine L. Westbrook (Chicago: ALA, 2004).

solutions to old problems without understanding sufficiently what they seek to replace. This is an unfortunate situation and it tends to exacerbate the divide between cultures, rather than encouraging conversation between these cultures. There are good reasons why libraries are slow to change, though clearly they have also been slow to articulate those reasons. Librarians can and should do better at recognizing the importance of what we bring to the table, and articulating its value to others.

What to Keep, What to Toss?

To avoid tossing out the baby with the bathwater, librarians should be able to take a look at traditional library values and practices, determining which we need to keep, and demonstrating our willingness to retire those that no longer serve our purposes. If we cannot move forward, legacy intact but eyes on the future, we risk further marginalization.

Behind the “MARC is dead” contention lies the idea that library catalogs are no longer useful or capable of providing access to digital materials.² This camp contends that the future depends on libraries abandoning their traditions and their catalogs, in favor of popular full-text search engines such as Google.³ This view, superficial at best, denies the reality that libraries provide more than access to current materials. Digital library consultant Karen Coyle, in a recent email on the Web4Lib discussion list, pointed out that because libraries pay attention to issues like longevity of access to materials, their goals are significantly different than commercial entities such as Google. As she

² Roy Tennant, “MARC Must Die,” *Library Journal* 127, no. 17 (Oct. 15, 2002): 26-28.

³ Google <http://www.google.com/> (8 Mar. 2008).

succinctly puts it: “Libraries occupy an information space that has a kind of geologic view of time.”⁴

If the view that Google can supplant library catalogs were the result of legitimate or empirical research into the value of search engines versus traditional access methods, rather than frustration with the pace of change in libraries, we might be able to discuss the merits of one approach or the other. As it is, most discussions of these issues rarely go beyond the superficial exchange of simple preferences, and the more troubling concerns that librarians express regarding privacy implications, fullness of coverage, and search sophistication beyond keywords are rarely addressed seriously outside of the library press.

Unfortunately, the failures of the current generation of library management systems, which have ignored the richness of MARC and refused to go beyond text-based strategies for searching and authority updating, has exacerbated these problems. Libraries using these clunky systems have been unable to take advantage of effective database management techniques to ensure that the collective investment in bibliographic and authority records, as well as publication patterns of serials, results in more effective results as well as management efficiency. These failures make Dick Miller’s XMLMARC strategies at Lane Medical Library at Stanford seem attractive to some, despite the fact that there is even less support in the XML world for the kinds of management strategies that libraries need to exploit their data effectively.⁵

⁴ Karen Coyle, [message in thread “Google Print as the Library’s Mission,” Web4Lib Electronic Discussion], <http://sunsite.berkeley.edu/Web4Lib/archive/0412/0197.html> (8 Mar. 2008).

⁵ Lane Medical Library, Stanford University Medical Center, “Medlane XMLMARC,” <http://xmlmarc.stanford.edu/> (8 Mar. 2008).

Is MARC Dead?

MARC, or more properly MARC 21, is the legacy metadata format developed by and for libraries. It was first developed in the mid 1960's at the Library of Congress (LC), partly to provide a basis for the production of the catalog cards which had been distributed by LC since the end of the nineteenth century. The original MARC format was developed for books, and later expanded to handle other formats, such as serials, maps, and computer files. The development and evolution of MARC into the rich schema available today has been possible because of the increasing experience and sophistication of librarians providing access to complex materials in new formats, and the willingness of these librarians to work together to invest in generalized solutions. MARC is a synthesis of the knowledge of these librarians, many of them domain specialists, who understood the value of collective effort.

The native syntax of MARC 21 is an elderly structure based on the original tape distribution needs of the 1960s when MARC was invented. Although still used to distribute MARC 21 records within library management systems and bibliographic utilities, outside those boundaries MARC records are increasingly distributed using eXtensible Markup Language (XML). Some years ago many technical services librarians were bedeviled by questions from the technologically ill-informed who became convinced that XML would replace MARC. However, since XML is only a syntax amounting to a generic package, requiring the addition of labeling and definition to convey meaning, the notion of "replacement" is problematic and ultimately a red herring.

In most cases, those who are most vehement about the actual or impending death of MARC have little notion of what MARC really represents, and what value it has in the

context of a more competitive metadata world. On one hand, MARC is a rich collection of ideas about organizing information, evolving from the card catalog era and the need to present enough data on a 3 x 5 card to enable a user to make decisions about whether to look for an item in the stacks. Since then, freed from the limitations of that 3 x 5 card and bolstered by the power of computers, MARC has enabled libraries to provide access to a great diversity of materials, conveniently dividing the information needed for this service into fields, indicators, subfields, and codes to enable manipulation by computers for the purposes of access and display. It is, admittedly, a somewhat ungainly beast, not simple to learn or apply, but it has supported the library world well for close to forty years.

MARC Alternatives?

Those who believe that MARC should be replaced by other, simpler alternatives have dismissed MARC as a dinosaur of an earlier age, citing its age, complexity, and numeric tags. On one level, it's difficult to credit some of these criticisms: there is, after all, an XML schema available for MARC, no fields are mandatory in all cases, and numeric tags make internationalization easier. Frequently cited alternatives, such as the Metadata Object Description Schema (MODS), LC's lightweight version of MARC 21 encoded in XML seem to ignore the fact that the terms are almost entirely defined by their relationship with MARC 21.⁶

Dublin Core, another alternative frequently cited in the search for a simpler alternative, has the advantage of an abstract model supporting its notion of describing

⁶ The Library of Congress, "Metadata Object Description Schema (MODS): Outline of Elements and Attributes in MODS Version 3.0," <http://www.loc.gov/standards/mods/v3/mods-3-0-outline.html> (8 Mar. 2008).

one resource at a time, while linking related descriptions (of creators, for instance).⁷ Interestingly, this model is similar to the one in MARC, in which text strings (or numeric values, in some more sophisticated linked systems) refer from unique values in bibliographic records to preferred terms in authority records. Breaking down names into parseable pieces and including affiliation with the name in the context of a bibliographic description, as MODS does, seems a questionable practice in a world moving inexorably toward explicit identification of linked descriptions with Uniform Resource Identifiers.

But perhaps the most troubling aspect of the marginalization of MARC is the tendency to deny the value of two of its most important legacies: the community-based process that brought together domain experts to develop the MARC formats as a standard data exchange useful for all, and the codified wisdom and experience of that community, reflected in the MARC documentation. The community that developed MARC, called the MARC Advisory Committee, includes the American Library Association's (ALA) Machine-Readable Bibliographic Information (MARBI) committee, US national libraries, the National Library of Canada and the National Library of Australia, the large bibliographic networks such as OCLC and RLIN, library associations such as the Music Library Association and Special Libraries Association, and library system vendors.⁸ MARC changes were shepherded by the Library of Congress Network Development and MARC Standards Office, and proposals were discussed online as well as at meetings during the semi-annual American Library Association conventions.

⁷ Dublin Core Metadata Initiative, "DCMI Abstract Model," <http://www.ukoln.ac.uk/metadata/dcmi/abstract-model/> (8 Nov. 2008).

⁸ Machine-Readable Bibliographic Information (MARBI), <http://www.loc.gov/marc/advisory.html> (8 Mar. 2008).

The community that grew around these meetings has provided much of the expertise behind many other metadata format developments. Several early and continuing members of the Dublin Core community got their start in MARBI, as did some of the leading lights behind VRA Core.⁹ The process of managing changes to the Dublin Core Metadata Element Set borrowed heavily from MARBI, in particular the emphasis on consensus and the reliance on stakeholders to review proposals.

MODS seems to be a relevant example of a new model of development, though certainly not the only one. MODS is not being developed by a broad community such as MARBI, nor is the process behind MODS development based on the successful MARBI model. Instead, the MODS home page cites “The Library of Congress’ Network Development and MARC Standards Office, with interested experts” as the developers of MODS and most of the review and discussion happens via a single mailing list, the Metadata Object Description Schema List.¹⁰ Discussions on this list tend towards the technical, not unusual in mailing list discussions in general, but not necessarily supportive of the growth of a “community” in the same sense of the one that drove MARC development. Most mailing lists consist of a large group of “lurkers” and a relatively small number of active participants, who set the tone of the discussion. Clearly, a list with primarily technical discussions is not likely to attract the participation of librarians without a strong technical background, no matter how deep their interest in metadata development.

⁹ Visual Resources Association, Data Standards Committee, “VRA Core Categories, Version 3.0,” <http://www.vraweb.org/resources/datastandards/vracore3/index.html> (8 Mar. 2008).

¹⁰ The Library of Congress, “Metadata Object Description Schema Official Web Site,” <http://www.loc.gov/standards/mods/> (8 Mar. 2008)

It may be that the model of development that served MARC so well is too expensive to replicate. The VRA Core used a similar model through version 3.0, but has now contracted with a single individual to manage further revisions, with review by the larger community. There are implications in these trends for the larger library community, both as new developments like MODS and VRA Core emerge, but also as the profession renews itself. As baby boomer librarians retire—and they're a particularly active cohort in libraries-- a development model that relies totally on virtual, rather than face-to-face interactions may well hinder the replacement of the graying heads with younger ones. The virtual community is a wonderful thing, but it favors the concrete technical discussion over the philosophical or experiential; the writer over the talker. Without any face-to-face contact, or any real-time, synchronous context to balance the sometimes dauntingly technical asynchronous one, such as the one formed now around the MODS mailing list, participation by the library community will continue to be sparse. We all lose in this scenario.

It is probably not coincidental that as MODS and METS have moved to the forefront of LC's development priorities, MARBI meetings focus on fewer and less critical issues, to the point that even a disinterested observer might wonder why so many people still gather together twice a year at each ALA meeting. Some of us still attend in hopes of hearing the stimulating discussions of the past, but we are generally disappointed. It's difficult not to wonder why MARBI was ignored as MODS and METS were being developed, rather than being used to effectively address community needs in the new metadata world.

But there are glimmers of hope, from surprising quarters. Thomas Delsey, a consultant specializing in information modeling, and known for his work on the

Functional Requirements for Bibliographic Records (FRBR), was appointed in September of 2004 as editor of the next edition of the Anglo-American Cataloging Rules (AACR3).¹¹ Interest in FRBR has spread from traditional libraries into digital libraries, and the discussion remains broadly useful, particularly as there is no syntax associated with FRBR to excite the technical cohort. The appointment of Delsey seems to presage (at least to the optimists among us) a possibly broadening role for the venerable cataloging rules.¹²

Clearly, the idea that a rejuvenated AACR3 might provide some welcome underpinnings for broader resource description has not escaped those involved with planning for AACR3. In a recent presentation, Barbara Tillett, chief of the Cataloging Policy and Support Office at the Library of Congress, stated several goals for the new development, including that they be compatible with other standards for resource description and retrieval, and be used beyond the library community.¹³ If this goal is to be met, attempts must be made to involve non-library stakeholders in the process as early as practicable.

Getting it Right

In October of 2004, I attended a panel presentation where three experts were asked to inform library practitioners by providing “evaluation” information about a large Dublin Core-based metadata repository. It was, in a small way, yet another version of the blind men and the elephant. The first presenter provided large numbers of

¹¹ Joint Steering Committee for the Revision of Anglo-American Cataloguing Rules, “Tom Delsey appointed as AACR3 Editor,” <http://www.collectionscanada.ca/jsc/aacr3editor.html> (8 Mar. 2008)

¹² Barbara Tillett, “AACR3: Resource Description and Access,” <http://library2.ust.hk/info/other/feb2005/aacr3pptjan2005-jak-hk-11feb.ppt> (8 Mar. 2008)

¹³ International Federation of Library Associations and Institutions, “Functional Requirements for Bibliographic Records Final Report,” <http://www.ifla.org/VII/s13/frbr/frbr.pdf> (8 Mar. 2008)

tables giving simple numeric tallies of elements used in the repository, with no more analysis than a relational database might reveal in ten minutes. The second provided results of a research project where users were carefully observed and questioned about what information they used when making decisions about what they were given in search results—i.e. useful data, and a good start on determining the usefulness of metadata, but with no attention paid to the metadata that was used behind the scenes, well before any user display was generated. The third presenter, a young computer scientist, relied almost entirely on tools developed for textual indexing, and, concluding that the diversity of the metadata was a problem, suggested that the leaders of the project should insist that all data providers follow stricter standards.

These presentations seemed sadly reflective of most attempts to approach the problems of creating and sharing metadata in the world beyond MARC. Traditional libraries built a strong culture of metadata sharing and an enormous shared investment in training and documentation around the MARC standard. The MARC development process codified the body of knowledge and practice that supported this culture of sharing and collaboration, building, in the process, a community of metadata experts who took their expertise into a number of specialized domains. We are now at a critical juncture. Moving forward in both realms, traditional and “new” metadata requires that we understand clearly where we have been and what has been the basis for our past success. To do that we need much better research and evaluation of our legacy and current models, a clearer articulation of short term and long term goals, and a strategy for attaining those goals that is openly endorsed and supported by stakeholders in the library community.

Conclusion

I remain eternally grateful for the opportunity to participate in MARBI and learn from other members and liaisons for well over a decade. As I have moved beyond the traditional library community and looked back, I have felt even more strongly that any future path for metadata standards development that disregards the values of such a group, as the venue for important discussion and directions, as well as the place where others interested in the important issues can learn, is risky at best. If the library community is not to be involved in the important developments that take librarians beyond MARC, what and who will drive future integrative thinking in the library world?

As many have noted, librarianship is a “greying” profession, with leaders who are now in their fifties and sixties. Preserving the important experience and knowledge held by those retiring in the next decade or so, while supporting the revolution and evolution that the computer age has enabled, requires that we pay careful attention not only to the product, but also the manner of building it. By understanding our legacy, and carrying our important values into the future, we honor those who taught us, and perhaps enable those behind us to benefit from our experience.