Present at the Creation
By Diane Hillmann

The Present Transition

Catching up with my reading recently I took a closer look at something I’d bookmarked, the RLG Programs Descriptive Metadata Practices Survey. [1] Karen Smith-Yoshimura wrote a tantalizing blog posting about the survey prior to the results being available, so given my interests in the subject I kept an eye out for the report. [2]

The report provides an important snapshot of the state of library use of metadata tools for the creation of metadata, the standards being used, and how well libraries are integrating the tools, standards and exposure of their metadata outside the institution of origin. The report pulls together the insights gained from the survey very effectively and makes some important points about how libraries see their mission and how this affects their processes of metadata creation and use of metadata tools.

The report discusses at some length findings that reveal:
* Libraries still tend to be focused on their “local users” even as it becomes more clear that the focus does not go the other way
* Libraries are generally reluctant to share their metadata via OAI or through web crawlers (this reluctance extends from MARC to other formats specialized for particular resources)
* Metadata creation tools are most often customized for a particular institution or particular group within an institution
* Use of controlled vocabularies is not optimized for sharing

The Continuing Myth of Local Users

In discussing the disconnect between library missions and new realities, the report points out that OCLC’s surveys of library users clearly reveal that users do not look first at library web sites for information about resources. Instead, they start at the general tools and search engines, so this lag in providing metadata to those services is clearly hurting libraries in their effort to engage their users with library services.

The intended audiences for locally created metadata correlate with the type of institution. In general, 80% or more of respondents serve an affiliated population (students, faculty, visiting researchers, and academic staff) but even more also cite the need to serve the “interested public.” More than half of all respondents identified a “primary audience,” which we would not expect in a networked world. This tendency to “look inward” may be a factor in the degree to which institutions expose their metadata for use by others outside their local population.

The question becomes, if geographically “local” users are using the general tools to identify resources--even physical resources--and using the catalog not for discovery of resources but discovery of shelf location for those materials not available digitally, what are the implications for those charged with creating metadata? Traditional cataloging practice emphasizes the “fit” of a catalog record within the context of a local catalog and local collection, but in the age of
WorldCat Local and mass digitization of books, this no longer makes sense. When (not if) the focus shifts to metadata optimized to external discovery mechanisms, we will likely see a great number of changes in practice--these are already anticipated by forward thinking administrators, and dreaded or denied by far too many catalogers.

**Metadata Tools**

Because the survey went not just to libraries, but also to museums and archives, the results expand considerably beyond the still MARC-centric metadata practices of libraries. The survey revealed a large amount of customization of tools, with few optimized for sharing either the tool, or the resulting metadata.

Although we saw some expected variations in practice across libraries, archives and museums, we were struck by the high levels of customization and local tool development, the limited extent to which tools and practices are, or can be, shared (both within and across institutions), the lack of confidence institutions have in the effectiveness of their tools, and the disconnect between their interest in creating metadata to serve their primary audiences and the inability to serve that audience within the most commonly used discovery systems (such as Google, Yahoo, etc.). [3]

The report asks some good questions about the possible reasons behind the more troubling of the survey results, but comes to no real conclusions about why tool development has lagged so far behind the need. One possible reason might be the lack of consensus on metadata standards, another might be the already noted disconnect between mission and audience. There is some brief comment in the report about how many respondents used automated tools, but no details about what tools those libraries were using and what their experiences had been. Also conspicuously lacking was any analysis of whether any of the tools blended automated and handcrafted approaches. This is an important gap, given that the economics of metadata creation are always cited when questions of quantity and quality come into play.

**Sharing Metadata Beyond the Library**

Among the most important insights from the survey concerns the surveyed institutions’ still limited view of their mission for metadata sharing. This is a particular issue for those institutions or parts of institutions focused primarily on unique objects. Traditionally, sharing is seen as only useful when others can re-use the metadata itself, not as a marketing tool to drive potential users of the content to institutional portals or resources. Libraries, with their sharing mechanisms tied strongly to their ILS’s and bibliographic utilities, pay lip service to sharing (primarily within their community) but still believe that the reason for doing so is to enhance community efficiency in creating metadata for traditionally published materials duplicated broadly over the library landscape. The result of this is very limited sharing of metadata beyond the institution of origin, based on a limited view of the importance of such sharing.

A third of all respondents do not have MARC metadata; this includes a majority of the museum collections and Institutional Repository respondents. A majority of those that do create MARC metadata expose it, predominantly through a Z39.50 server; a quarter do so by using the Open Archives Initiative—Protocol for Metadata Harvesting (OAI-PMH). Only one respondent uses SRU (Search/Retrieval via URL) or SRW (Search/Retrieval via the Web). About a quarter of the different workplace environments do not expose their MARC metadata. [1]
To a great extent, this resistance to sharing may reflect the incomplete transition to an integrated view of library services as part of the “platform of the Web.” It may also be the case that the success of many libraries in bringing bodies into their physical buildings has blinded them to the differences in the purpose of those visits—many are surely responding to the attempts of libraries to become social spaces as well as research destinations. Regardless of the causes, the result of this blindness to metadata’s marketing potential tends to marginalize libraries as players in the places where their users work, even as they make prodigious efforts to digitize their holdings to make them more broadly available. The report makes a very strong point that this behavior needs to change:

While most count the public among the audience for their resources, respondents still see their primary audience as restricted to affiliated users (students, faculty, and staff). Arguably, both affiliated and unaffiliated audiences congregate in large-scale information hubs, which current disclosure strategies target only to a limited degree.

To reach users wherever they are, we as a community need to disclose more metadata to OAI harvesters, Web crawlers, and also push metadata out directly to information hubs. For disclosure to be effective, search engine optimization is crucial. [1]

It should also be recognized that “sharing” via Z39.50 is by definition only sharing with other libraries, and only with MARC data. Use of OAI-PMH represents a different kind of sharing, not designed to represent someone else’s data within the context of a library catalog, as Z39.50 does, but sharing data without reference to its ultimate use. While it’s absolutely true that some OAI-PMH data comes with Creative Commons license restrictions against commercial use of the data, non-commercial uses can potentially be quite varied.

**Controlled Vocabularies**

Some of the most interesting results of the survey are not that the majority of controlled vocabulary use involves Library of Congress Subject Headings (LCSH), the Library of Congress Name Authority File (LCNAF) and the Art and Architecture Thesaurus (AAT)—this we could have predicted. But the report goes on to note that:

Nevertheless, about half the respondents build and maintain one or more local thesauri. Segmenting responses by workplace environment indicates that museum collections and digital libraries build and maintain local thesauri the most, and libraries do so the least. The types of information included in these local thesauri, in order of the response rate: genres of materials, topics, people and organization names, places, and time periods. Since a large majority of all respondents agreed or strongly agreed that “user-supplied tagging in addition to controlled vocabulary is the best for the resources we describe” and an even larger majority agreed or strongly agreed that “it is critical to provide controlled vocabulary to the resources we describe,” the relatively frequent use of local thesauri may suggest a need to make it easier to contribute to shared terminologies. [1]

It seems clear from these responses that despite the fact that libraries and librarians are well known for their support for and use of vocabularies, they prefer that these
vocabularies be developed and maintained by somebody else. This strategy has worked reasonably well for a long time, and relying on others to manage the vocabularies has significant short-term economic benefits for individual institutions, but long term it has created dependencies that are hampering the community as it seeks paths to the future.

The recent report from the Library of Congress Working Group on the Future of Bibliographic Control made a strong statement in support of the necessity for new ways of thinking about meeting the community's needs:

_The future of bibliographic control will be collaborative, decentralized, international in scope, and Web-based. Its realization will occur in cooperation with the private sector, and with the active collaboration of library users. Data will be gathered from multiple sources; change will happen quickly; and bibliographic control will be dynamic, not static. The underlying technology that makes this future possible and necessary—the World Wide Web—is now almost two decades old. Libraries must continue the transition to this future without delay in order to retain their significance as information providers._ [4]

Collaboration, decentralization, internationalization, with the web as platform—it can't come soon enough, in my view.


