

**Association for Library Collections and Technical Services (ALCTS),
a division of the American Library Association
Written Testimony for the Library of Congress Working Group on the Future of
Bibliographic Control
Second Public Meeting: Standards and Structures,
Chicago, IL, May 9, 2007**

The Association for Library Collections and Technical Services (ALCTS) appreciates the opportunity to provide written testimony for the Working Group's consideration, in advance of its upcoming public meeting in Chicago. A paper that responds to the reports of the Working Group's first meeting on the topic of "Users and Uses" will be forthcoming. This written testimony addresses the five questions asked by the Working Group in its background paper.

Q1. What kinds of standards and structures are needed to provide effective bibliographic control in the environmental spectrum spanning consumer uses and management uses? How can we make better use of current standards and structures in meeting both consumer and management user needs? What relevant communities need to have input and what organizational structures would best support this?

Standards and structures are needed that are global in terms of world-wide applicability (e.g., useful in scores of languages besides English) and in terms of user groups. This is a situation faced not only by libraries, but by multiple sectors, such as business, government, and academia. The way forward will mean collaborations among all parties in the information chain: publishers, governments, aggregators, authors and other creators of content, readers and other users of information, computer/information professionals, and others.

The enormous quantity of metadata based on current standards (AACR2, MARC, LCSH) has greater potential value than ever, particularly given the ease with which it can be easily "mined" for new applications. The consistency of use of those standards that characterized cataloging practice for the last half of the twentieth century is already proving to be of great benefit as our services develop. Briefly mentioned examples include OCLC's FictionFinder; Columbia University's HILCC (Hierarchical Interface to LC Classification), recently acquired by Serials Solutions; North Carolina State University's well-known application of Endeca for faceted searching. In the realm of social networking, LibraryThing's use of MARC record data and its implementation of LCSH is significant. This point should not be lost as our existing standards evolve and new standards become mainstream.

There is a movement away from reliance on ILS vendors and proprietary software for major enhancements to the "user experience." Nevertheless, the investments made by libraries in ILS systems will continue to be significant, and the systems themselves are the primary points for storage and manipulation of by far the greatest proportion of core metadata (bibliographic, financial, vendor and end user). To reinvigorate the library/vendor relationship, we suggest a new approach to system development. It is possible that a leadership group, drawn with care from across the entire spectrum of librarianship, could envision a "new ILS" with guaranteed basic functionalities, both of traditional types and of those needed to serve present and foreseeable future generations of end users. This set of guaranteed functionalities would become a basic set of specifications. At the same time, there would be, in every library/vendor contract, the guarantee of far greater fluidity of the ease with which traditional core functions may be enhanced, and the ability to link proprietary systems with populist emerging technologies. To put it another way, it is time for the concept "hard-coded" to become a thing of the past.

With regard to community input, it is essential that a broad spectrum of user constituencies is actively included as a normative practice. This spectrum needs to go well beyond academic (particularly research)

and government libraries, and the larger players in the information industry. Public, school and special library representatives should be centrally involved: leadership groups, which may of course be fuzzily bounded and informal, should represent the population at large. "User communities" should also be understood to include linguistic and cultural communities, not only professional communities defined by constituency. To allow for optimal participation, grassroots settings are required, in addition to national conferences and by-invitation meetings. Forums could include state and regional meetings, facilitated by state library associations, OCLC regional service providers, regional resource-sharing consortia, LIS schools, and so on. These can complement online activities, including the uses of wikis, blogs and online forums, virtual meetings, and collaborative development of open-source software tested using significantly large datasets.

Q2. Libraries and related cultural heritage organizations have made a major investment in controlled data. These include structures for organizing subjects, personal and corporate names, place names, roles and relationships, time periods, etc. What role will this data play in networked environments? What is its relationship to the semantic web, tagging, or other newer approaches? How does this data work across database silos? How are supporting infrastructure pieces (gazetteers, controlled vocabularies, etc.) situated and maintained?

It has become evident that, to effectively address these questions, a significant investment is required in interoperability at all levels. Many demonstration projects, as well as working implementations of limited scope, have shown the potential benefits of search, retrieval and display in environments where the user is not required to understand the idiosyncrasies of a given "silo." What is required is a move toward implementation of the best practices and effective outcomes already discovered, in large-scale end-user environments. This in turn means acceptance of the "perpetual beta" ethos now characteristic of new product development and satisfying user experiences. It is presumed, of course, that further research and experimentation will continue to be essential, but the time is now to invest in production environments, beyond pilot projects and short-term specially-funded efforts.

Types of interoperability include, at least, among languages and scripts, among data structures, and among user community or disciplinary vocabularies. The latter involves building bridges among authority-controlled metadata systems per se, and between these systems and loosely- or unstructured end-user metadata such as tags. We would do well, in the United States, to learn from effective interoperability projects developed in areas of the world where multilingual discovery is an ordinary need. (An example in the area of subject metadata is the MACS project, now hosted at Brussels Free University.) The concept of interoperability may be extended to include combinations or "mashups" of multiple types of data: for instance, place names in all of the world's languages with geospatial coordinates attached to those names.

The question of how supporting infrastructure elements are "situated and maintained" can be thought of by using the image of interlocking decentralized systems. Assuming the development of multiple levels of interoperability as a basic element of information retrieval systems, it should be possible for a greater variety of controlled data systems to be used in a common environment. In the case of subject metadata, for instance, there would be no question of "shoehorning" the subject vocabulary of a given community – whether that vocabulary is highly or slightly structured – into the constraints which are a necessary part of mainstream systems. The latter (such as LCSH, LCC and DDC) would not only continue to play major roles, but would likely gain in value when flexibly implemented in a variety of environments.

The expertise that librarians have developed in authority control is one of the most important elements that libraries have to offer to the larger information environment. The value of it is recognized by popular information sources, such as the Internet Movie Database and LibraryThing. We need to develop systems that enable authority control to work more effectively in multiple information environments, as well as to

make the underlying concepts more accessible for end users. Developing interoperability between controlled vocabularies and user-provided metadata may be crucial, not only for helping users make effective use of controlled metadata, but also in terms of user assistance and understanding.

The provision of authority control is by nature a complex process. As global interoperability between and among databases of controlled data becomes the norm, the value of this investment in trained human intellect will become increasingly evident. Nevertheless, there are barriers to increased effectiveness and efficiency in the practice of authority control which should be addressed. For example, NACO trainers state that the requirement to consult multiple sources in training (e.g., AACR2, LCRI, DCM, NACO manuals, rule summaries on the PCC NACO pages, etc.) presents as much difficulty, or more, than the complexity of the rules themselves. Other obstacles are presented by factors such as highly complex rule interpretations and apparently contradictory conventions for construction of qualifiers and uniform titles. On another level, there is a need for more effective sharing of authority-controlled metadata, beyond local creation and maintenance of individual databases. Just as the mechanisms which enable interoperability need not be replicated in every local system, so too should local systems be able to make direct use of central, international databases while still allowing for local needs. This holds the potential for a significant reduction in redundant effort.

Q3. Data is created to be processed by applications. We mine it for meaning; merge and manipulate it for display; use it to support supply chains and inventory control; share it between repositories and discovery environments. Are our standards and structures appropriate to this reality?

We would like to offer, to begin with, a caution in response to this question. Data is, in the first instance, created by human beings, directly or via mediated techniques, so that others may gain information and understanding through interaction with that data. (For example, the data underlying works of literary value is not created by authors for the primary purpose of being scanned by an automatic character recognition program.) However, when human beings create data in an information sharing environment pervasively mediated by computers in communication networks, then it is essential to create the data they wish to share with one another in forms that are easily and richly processed by computerized methods. It is in this sense that data is created to be processed by applications.

It is certain that the standards and structures which presently exist are not adequate to support many of the applications desirable and needed for sophisticated data processing in the service of multiple end-user tasks. This is true for the two generalized user environments described by the Working Group as “consumer” and “management,” as well as those of identifiable user groups (e.g. parents of young children, community college students, information industry professionals). This brings us again to enhanced interoperability. It is unlikely that overarching “rules, guidelines, models, and structural schema” will be developed which enable the widest variety of applications to share and process data in the many ways described in the above question. It may not even be desirable to develop such universally controlling mechanisms. What is more likely is a continuation of what we have seen for decades, as described in the saying, “The great thing about standards is that there are so many of them.” In this connection, our attention should continue to be placed on enabling standards and structures, those existing as well as to come, to interoperate flexibly.

Q4. What requirements are placed on our bibliographic structures through new application areas, such as mass digitization and greater off-site storage, or the desire to create richer user interfaces and integrated discovery environments?

The primary requirements are elegance and fluidity. Elegance implies simplicity, in the sense of facilitating communication among applications, not in the sense of losing richness of metadata content or granularity in coding. Fluidity implies the ability to translate metadata, with its associated coding, into

multiple interfaces and environments. Elegance and simplicity together imply a clean separation of the intended functions of different metadata elements and types from display conventions. At the same time, metadata is associated with context, or provenance, which must be preserved as it is frequently key to the meaning of a given term (e.g., subject terms established according to LCSH as compared with MeSH). Metadata from multiple sources, whether stringently or loosely controlled, should not lose differentiation even when usefully residing in the same system.

Q5. Libraries now manage different flows of data, created within different regimes, much of it outside the library environment. They also want their data and services to appear in other environments. At the same time, we see more reuse and flow of data across publishers, libraries, agents, other bibliographic services, etc. What does this mean for our bibliographic standards and structures?

This question recapitulates the general themes discussed. Standards and structures need to enable metadata drawn from different disciplines and communities to be reused intelligently in a variety of end-user environments. “Library-created” metadata, whether of established or recently developed types, should be able to flow easily into applications created outside librarianship proper. Similarly, metadata types with provenance outside librarianship should be able to be incorporated into whatever “library systems” become, for the purposes of building on the strengths of library-created metadata as well as addressing its weaknesses. Implied are the development of standards and structures which allow fluid sharing and mixing of data types, preservation of context (the metadata’s “original intelligence”), granularity in indexing and display to any degree desired, decoupling of markup from display, and translation/transformation into the conventions of multiple end-user environments. It is important that standards of types such as RDA and MARC be further developed to enable “hooks” to multiple other standards. At the same time, these standards need to be supplemented by others whose sole function is to serve as multiplug adapters, in a sense. Analogously, “structures” or organizational bodies such as the JSC, MARBI or PCC would want to consider their standards-developing activities in this light.

References

HILTT: <http://www.columbia.edu/cu/libraries/inside/projects/metadata/hilcc/>

LibraryThing: <http://www.librarything.com/>. Note that the “book suggestion” engine provides suggested reading of interest based on several criteria, including “similar library subjects and classifications” drawn from MARC records.

MACS: <https://macs.vub.ac.be/pub/>

NCSU: <http://www.lib.ncsu.edu/catalog/>

OCLC FictionFinder: <http://www.oclc.org/research/projects/frbr/fictionfinder.htm>

Prepared by the ALCTS Cataloging and Classification Section on behalf of the ALCTS Board of Directors.

Submitted to the LC Working Group by the ALCTS Board of Directors.
April 24, 2007