

Working Group on the Future of Bibliographic Control
Economics and Organization of Bibliographic Data
July 9, 2007
Meeting Summary

Nancy J. Fallgren, Johns Hopkins University
Consultant to the Working Group

The third public meeting of the Working Group on the Future of Bibliographic Control concerned the economics and organization of bibliographic data. While bibliographic control is a service that provides a public good, it does generate costs which are borne by the institutions that provide it. Deanna Marcum, Associate Librarian for Library Services at the Library of Congress, underscored this point in her opening remarks. The Library of Congress is involved in a broad strategic planning initiative to provide services more efficiently and streamline costs, while at the same time remaining cognizant that its actions can have repercussions in the wider library community.

The speakers invited to the meeting were drawn from a wide variety of libraries and related commercial organizations (the speaker agenda is available at <http://www.loc.gov/bibliographic-future/meetings/agenda-july9.html>) in order to gain a broad perspective of economic and organizational issues concerning bibliographic control. In addition, comments from the public were actively solicited to augment the prepared testimony.

Setting the stage for the meeting, Rick Lugg, a library consultant, talked about cost-benefit ratios and noted the need to consider the total cost of bibliographic control throughout the lifecycle of a resource, from selection to digitalization or withdrawal. He cited current practices in bibliographic control that generate costs (both monetary and non-monetary):

- Redundant or unnecessary workflows
- Creating original MARC records
- MARC format
- Metadata and database maintenance
- Managing backlogs
- Lost opportunities to use cataloging expertise to greater benefit

Redundant, duplicative, and unnecessary workflows all are indications of inefficiency in the bibliographic control process. They include a variety of common cataloging practices such as reviewing and/or editing bibliographic records created by the Library of Congress and Program for Cooperative Cataloging (PCC), seeking perfection in creating original cataloging records, creating unique call numbers for every item, and not making use of pre-existing metadata from other sources. In general, there was consensus among the speakers that these inefficiencies can be combated by not seeking perfection in bibliographic records. In some cases, that may mean accepting bibliographic records created by other stakeholders without reviewing them or editing them for minor errors, such as punctuation. For example, there is a dual standard in place that expects perfection when creating original cataloging records, but accepts batch loaded vendor records without quality review. Mary Catherine Little (Public Libraries) stated that for popular materials in public libraries, efficiency may mean accepting records at less than full cataloging

levels based on the need for expediency in making materials available, the expected lifespan of an item, and the sufficiency of the existing metadata to enable discovery.

Several speakers suggested that the current mindset of creating a perfect record that need never be edited again is an inefficient and unrealistic expectation. Karen Calhoun (OCLC) described metadata as having a lifecycle, during which enhancements are incorporated over time. Bob Wolven later characterized this lifecycle as a fluid, rather than linear, process in that we cannot know at the start what metadata will be important at a later stage.

The cost of original cataloging was exemplified by several speakers' descriptions of business models with common basic features: one entity bears the cost of creating original metadata in the form of bibliographic records, authority records, or taxonomies (per Rick Lugg, creation of one bibliographic record for a monograph can cost between \$150-\$200), the metadata is made available for others to use, but the creator of the original metadata does not recoup the full cost of its labor from the other organizations that are sharing the benefits. The financial inequality of this practice and need for a new, more equitable business model was pointed out by speakers representing both libraries and commercial organizations. Susan Fifer Canby (Special Libraries) mentioned that the National Geographic Society library is becoming more involved in e-commerce, which may be one means of cost recovery for creating original metadata on marketable resources, such as photographs.

Despite decrying the business models currently in place, it was acknowledged that reducing costs via the practice of sharing bibliographic metadata and relying on the work of other stakeholders are means by which libraries keep cataloging budgets under control and maintain some efficiency. Many speakers advocated increased collaboration and partnerships as a means of becoming more efficient. However, the question of how to share the cost of creating the original metadata more equitably remains largely unresolved.

Several speakers cited the MARC format as inefficient, often in that it is inadequate for their needs. Rick Lugg remarked that the complexity of the MARC record, specifically redundancy between the fixed and variable fields, creates a barrier to efficiency among catalogers. Lizanne Payne (Library Consortia) noted that, as an inventory system, MARC format does not easily handle identification of the physical location of materials in shared storage facilities, where items are often shelved by size, rather than call number. Susan Fifer Canby observed that much of the material held by special libraries is not necessarily amenable to description in MARC format.

The handling of MARC records in various integrated library systems (ILS) also was described as an inefficient process. Rick Lugg noted inefficiencies in the dual role of MARC records as both a transaction vehicle for inventory control and a tool for discovery. As a transaction vehicle, a minimal MARC record is created for later matching to and replacement by a more complete, complex record for discovery. In its discovery role, the master MARC record is encumbered by linked records for items and holdings, making record maintenance more difficult to manage and creating more laborious database maintenance, particularly in deleting a MARC record, along with all its appropriate linked records, from an ILS.

In a similar vein, Mary Catherine Little noted that every time a MARC record is “touched,” it creates a cost. She described a common process where vendors supply MARC records to a library for download into its ILS. Each proprietary ILS breaks down the MARC records to accommodate its MARC display format and to attach linked holdings, item, and authority records to the master MARC record. She suggested that this process creates unnecessary ILS vendor overhead and database maintenance overhead. For library consortia, Lizanne Payne described the need to create three levels of records for one manifestation: a master manifestation level record for the entire consortium, a local level record (where members may exercise more precise subject analysis), and individual member holdings records. Both Little and Payne commented that implementing FRBR would help to alleviate these duplicative processes.

Cataloging backlogs are created by a number of factors including ILS conversions, inefficiencies in the cataloging process, and the sheer volume of materials being acquired. Backlogs of print materials are visible and, therefore, demand attention; however, digital backlogs are invisible, so that their extent is often unknown and easier to ignore. Backlogs produce costs in the labor spent managing them, in delaying resource availability to users, and in “opportunity costs.”

Rick Lugg defined opportunity costs as a question of setting priorities. Stated differently, it involves freeing human resources from routine tasks in order to focus on more value-creating opportunities. Lugg identified several areas of opportunity where cataloging expertise should be exploited more heavily, include cataloging unique materials (such as special collections and archives), institutional repositories, dissertations and theses, non-MARC metadata projects, course management systems, and mass digitization projects. The cost to libraries is in the lack of involvement, or delayed involvement, as a partner or leader in projects where cataloging skills are needed, such that the need is not well met or it is fulfilled with capabilities developed outside the library community.

In addition to enumerating the costs of bibliographic control, both the invited speakers and the public suggested initiatives that could be undertaken to increase efficiency in bibliographic control. These initiatives include broadening collaboration and partnerships, use of pre-existing metadata, repurposing library created metadata, structural reorganization, incentivizing creation of original metadata, using automation to assist catalogers, and revising and streamlining standards.

As Bob Wolven commented, the methods discussed to improve efficiency were largely collective. Michael Charbonneau (PCC and Large Research Libraries) talked about the success of PCC programs (particularly its role in creating authority and CONSER records), a voluntary international coalition of stakeholders that both create and use bibliographic records. The PCC seeks to continue its success in part by increasing and diversifying its membership, entering cooperative ventures with book vendors and publishers, and working with international agencies to align authority records. Linda Beebe (Abstracting and Indexing Services) stated that future success lay in dismantling silos and increased collaboration. Both Beacher Wiggins (Library of Congress) and Mary Catherine Little discussed their growing relationships with international book vendors to assist in cataloging foreign language materials. Little suggested the Library of Congress should develop relationships with foreign national libraries to share cataloging of

foreign language materials. Following a question concerning libraries that cannot afford to be OCLC members, Karen Calhoun replied that OCLC is looking into ways to make its services more broadly available and seeking partnerships with other organizations in the “supply chain.”

Capturing bibliographic data further upstream was discussed in various contexts. Data captured from publishers and passed through various value-adding processes by other parties in the environment could result in widespread gains. It was recognized that this could be achieved under various organizational and business models. Christopher Cole (National Agricultural Library - public testimony) noted that the National Agricultural Library (NAL) has been using basic metadata records from publishers as the basis for indexing records, to which librarians add quality in the form of access points. Cole stated that this process results in tremendous cost savings with no loss of quality.

Similarly, Dianne McCutcheon (National Library of Medicine - public testimony) explained that indexing personnel at the National Library of Medicine (NLM) do not see a positive cost-benefit ratio to name authority research because it is too laborious, while catalogers at NLM believe there is a positive cost-benefit. Recognizing the value of name authorities, she suggested that the records could be created more efficiently and by a broader community, if personal data created by the authors themselves could be collected by their publishers and made accessible to the library community.

Sources of pre-existing metadata need not be limited to publishers. Christopher Cole noted that existing sources of metadata, such as film metadata from authoritative websites, should be drawn upon in lieu of painstakingly re-creating the metadata from the item in hand, as required by current cataloging rules.

Christopher Cole also suggested that sourcing metadata should be a two-way street. The value of authority records and controlled vocabularies to the library community and beyond, along with the need to promote broader use of such metadata, was discussed in the two previous public meetings and reaffirmed here. Dianne McCutcheon and Christopher Cole offered concrete examples of repurposing controlled vocabularies beyond their original intent, including use by other communities. The NLM’s medical subject heading thesaurus (MeSH), created as an index for articles in the MEDLINE database, also is repurposed as a subject heading tool for cataloging in the medical library community. The NAL’s Agricultural Thesaurus is a valuable bibliographic tool for subject access that, per Cole, is not expensive to maintain. It also has sufficient value to international agricultural organizations outside the library community that they have offered assistance in NAL’s translation of the thesaurus to Spanish, thereby helping to defray the project’s cost.

Beacher Wiggins explained an initiative at the Library of Congress to create efficiency by restructuring the selection, acquisitions, and cataloging workflow. The goal is to supply as much descriptive cataloging as possible the first time an item is handled in the selection-acquisitions process, minimizing the number of times a record is “touched.” Technicians will have primary responsibility for descriptive cataloging, freeing professional catalogers to focus on authority control, subject analysis, classification, and digital resources.

Dianne McCutcheon specifically cited the need to incentivize organizations to create original metadata for sharing. McCutcheon explained that while OCLC offers a one time financial credit to contributing members for the creation and/or updating of a bibliographic record, there is no further financial benefit to the originating institution when the record is shared with other OCLC members. She suggested that in order to encourage creation of original bibliographic records, rather than waiting for another OCLC member institution to do the work, the originating institution should receive remuneration each time a record is used.

Both Mechael Charbonneau and Dianne McCutcheon suggested that advances in automation should be used to help catalogers be more efficient, not to replace them. For example, automated indexing programs can suggest index terms to catalogers, who then determine the appropriateness of the suggested terms.

Echoing topics discussed at the Working Group's previous meetings, several speakers advocated revisions in standards for bibliographic control, including cataloging rules and formats, as a means of creating efficiency. Susan Fifer Canby noted that special libraries in particular require a flexible metadata standard for published and unpublished content that may be unique to each organization. Christopher Cole suggested that cataloging rules should be more forward looking in expectation of change. Diane McCutcheon suggested that standards should be simpler, more flexible, more broadly applied, and developed more quickly. Bob Nardini (Vendors) suggested that there should be wider agreement on bibliographic standards in the library community.

Changes to the cataloging landscape will incur education costs on two fronts: educating users to use our systems to full advantage and educating librarians and paraprofessionals to adapt to a changing work environment. Linda Beebe pointed out that our users do not necessarily know how to use our systems to their best advantage. We should be cognizant that we may need to educate our users as to what they can do with our systems and we should not view simpler discovery tools as "dumbing down" the system.

Beacher Wiggins acknowledged that restructuring the workflows at the Library of Congress will require a broader group of staff to become familiar with cataloging practices, incurring training costs at least initially. If other suggestions are put into practice, catalogers will need to learn to use new technologies designed to assist them, become more conversant with a variety of metadata standards, and take their skills to emerging and unfamiliar arenas where their expertise is needed. As libraries struggle with a steady decline in professional catalogers, questions of finding sufficient professional catalogers to create original metadata and the role that library and information science programs should play in educating professionals for a changing landscape in bibliographic control also were raised.