



December 14, 2007

Olivia M. A. Madison
Dean of the Library
302 Parks Library
Iowa State University
Ames, IA 50011-2140

omadison@iastate.edu

Dear Olivia,

6565 Kilgour Place
Dublin, OH 43017-3395 USA

T +1-614-764-6000
1-800-848-5878
F +1-614-764-6096
E oclc@oclc.org

WWW.OCLC.ORG

Thank you for the opportunity to respond to the November 30 draft report and recommendations of the Library of Congress Working Group on the Future of Bibliographic Control. OCLC congratulates the Working Group (WG) on gathering community input widely, engaging in a thorough and detailed analysis of findings, and producing actionable recommendations.

In the “guiding principles” section of the report, the WG urges a redefinition of bibliographic control that embraces a wide array of materials, diverse user communities, myriad venues where information is sought, and many sources of metadata for organizing collections for discovery and use. OCLC too is striving to embrace and adapt to these new underlying conditions for bibliographic control while being true to its mission to reduce library costs and connect the world’s libraries. As you have urged in the draft report’s cover letter, we at OCLC stand ready to step forward and take responsibility for assisting the Library of Congress and all libraries with realizing the goals set out in the WG’s report.

Several OCLC initiatives speak directly to the WG recommendations. In general, OCLC Metadata Services is pursuing a program with two strategic thrusts: (1) innovation and cost reduction for libraries; and (2) the realization of a collaborative global metadata network.

Our intent is first to offer libraries bibliographic control services based on re-engineered, technology-based methods that lower library costs for collecting and organizing more mainstream publications, thus freeing up local staff resources for new initiatives and services of key strategic importance, such as making “hidden collections” more accessible. We are assembling a set of services based on mining WorldCat to optimize library-, publisher- and vendor-supplied metadata.

Second, our aim is to support the realization of a global metadata network for libraries, museums, archives, and other organizations that want to make their rich collections highly visible to a worldwide audience. Related to this goal, OCLC is developing services to bring a wide variety of metadata into WorldCat and create reliable linkages to the wide array of materials in the diverse, high quality collections that are under the stewardship of cultural heritage, scientific and technical organizations all over the world.

OCLC Programs and Research has developed a work agenda that promises to be of keen interest to those attempting to position library and cultural heritage organizations for the future. The goal of OCLC Programs and Research is preserving and renewing the value of research libraries, archives, and museums. A synopsis of the work agenda is available at <http://www.oclc.org/programs/ourwork/agenda.htm>. The leaders of OCLC Programs and Research are preparing a detailed response for the WG’s and LC’s consideration, which will come to you under separate cover.

Please find below OCLC’s enterprise-wide comments on specific WG recommendations relating to all of our activities and plans. For convenience I have gathered our comments into the five general areas in the WG report.

Efficiency of Bibliographic Production

1.1 Eliminate Redundancies

The WG report criticizes the customary library practice of making changes to records in local library systems but not sharing these changes with the broader community. Library managers quite rightly want their staff to work in one system (usually their own ILS cataloging client) and not double key changes into whatever shared cataloging system they use as a source of records (often OCLC). While OCLC's current policies and incentives for master record enhancement undoubtedly have some impact on library managers' decisions about cataloging workflows, the principal driver of this particular form of redundant effort is the necessity, up to now, of having thousands of local systems for carrying out bibliographic enrichment and maintenance. OCLC is undertaking an analysis of its record contribution and enhancement program, but we believe it is a mistake to expect even significant changes in OCLC's program to have much effect on managers' workflow decisions.

Until edits made in a local cataloging client can be effortlessly propagated to WorldCat, or some other sea change occurs in the array of tools customarily used in libraries to carry out bibliographic control, managers are likely to continue choosing the local catalog as their database of record. Along these lines, OCLC is experimenting with the SRU (Search and Retrieval with URL) update protocol to enable seamless metadata exchange and enrichment between WorldCat and the CBS union catalog software that is used in Dutch, Australian, French, German, and UK union catalogs. This allows automatic background update of WorldCat in real time at the same time as updating the regional group catalog. OCLC would be willing to collaborate with others to promote SRU update and similar standards.

The WG report also expresses concern about wasted resources that result from the presence of duplicate records in consortial databases or OCLC, and indeed this is a serious issue. By the end of FY 2008, OCLC Metadata Services intends to complete a business case for renovating the WorldCat metadata quality program. The plan will include proposals for reintroducing the Duplicate Detection and Resolution (DDR) program and for managing "allowable" duplicates (e.g., the same manifestation cataloged in French and English).

1.1.1.1 All: Be more flexible in accepting bibliographic data from others (e.g., publishers, foreign libraries) that do not conform precisely to U.S. library standards.

And

1.1.2.3 All: Use descriptive cataloging provided by book vendors and non-U.S. libraries whenever available.

Over the last three fiscal years, OCLC has energetically pursued a strategy to transform WorldCat into a networked information resource that has relevance and utility for libraries around the world. In August 2006, OCLC launched WorldCat.org, whose primary purpose is to heighten the visibility of the world's library collections on the Web and to drive end-user searches from the open Web to group and local library catalogs.

To make WorldCat.org relevant and useful in a global context, OCLC has sought agreements with non-U.S. national libraries and large consortia. One result of loading so many large non-U.S. data files is that today, 47% of the records in WorldCat describe materials in languages other than English.

OCLC processed about 22 million records from non-U.S. libraries over the last three years. In the coming year, we expect to process over 72 million records from outside the U.S. Some recent or pending loads include the Bavarian State Library, two large consortia in Germany representing hundreds of libraries, Glasgow University, the Jewish National and University Library, the National Library of Australia, the National Library of New Zealand, the National Library of Sweden, the National Central Library of Taiwan, National Taiwan University, the National Library of China, and the Brazilian Institute of Information in Science and Technology.

To keep up with significantly increased demand for batchloading from non-U.S. libraries, as well as to optimize the usefulness of WorldCat.org for U.S. libraries, OCLC has launched a major development effort to re-engineer the WorldCat data ingest processes. Soon, WorldCat will be able to automatically accept a broader range of

metadata from various sources resulting in quicker and broader visibility of non-U.S. library collections in WorldCat.

With a growing presence of non-U.S. descriptive cataloging, heading forms, and subject schemes in WorldCat, OCLC is taking a broad perspective of what metadata is necessary and important to serve libraries and end users both inside and outside the U.S.

1.1.1.3 All: Develop standard crosswalks for the conversion of vendor data to library system formats.

Over the past several years, OCLC has developed crosswalks from non-MARC data and continues to enhance these crosswalks, especially in the areas of publisher ONIX data and formats used to describe digital objects. OCLC allows use of the crosswalks as a web service and also plans to greatly expand the use of crosswalks in ingesting metadata into WorldCat.

1.1.1.4 All: Develop managed processes for creating and sharing conversion programs so that programming is not done redundantly at multiple institutions.

The OCLC crosswalk web service and the addition of records to WorldCat using crosswalks will benefit multiple libraries in the conversion and use of metadata that originates in non-MARC formats. We're exploring, for example, the development of crosswalks for metadata standards used in the museum and archive communities. The conversion programs are being developed using an architecture that could be licensed for use by third parties.

1.1.1.5 All: Work with resource providers to coordinate data sharing in a way that works well for all partners.

OCLC continues to expand the number and scope of supply chain partners that are contributing to WorldCat and interfacing with WorldCat for record delivery to libraries. For example, we currently partner with over fifty materials vendors to reduce the cost of cataloging for member libraries.

In addition, a new pilot program—called for now “Next Generation Cataloging”—with publishers and other supply chain partners will greatly expand the effort to coordinate data sharing and publisher/supply chain metadata re-use.

1.1.1.6 All: Demonstrate to publishers the business advantages of supplying complete and accurate metadata.

Through the Next Generation Cataloging project, OCLC is developing strong relationships with book industry standards organizations such as the Book Industry Study Group (BISG) and is reaching out to content creators and providers to assist in the production of quality upstream data that can be effectively used in multiple contexts.

1.1.3 Automate the CIP process

OCLC proposes that LC and OCLC take this on as a joint initiative. LC should be able to benefit from OCLC development in the context of our Next Generation Cataloging project, which will feature automated conversion, ingest and enhancement of ONIX metadata. OCLC could share techniques and best practices and/or directly provide converted and enhanced metadata to LC.

1.1.4 Re-Examine the Current Economic Model for Data Sharing in the Networked Environment

OCLC plans a symposium on this issue in connection with the Next Generation Cataloging pilot program. Leaders from libraries, publishers and other content providers and content supply chains will be included. A tentative date for the symposium is ALA annual, June 2008.

1.2 Increase Distribution of Responsibility for Bibliographic Record Production

We agree that long-term reliance on LC provision of cataloging records is gradually becoming less sustainable. In some respects however, the WG report may understate the options available to small libraries, should LC curtail its bibliographic control activities to some degree. For example, OCLC CatExpress subscriptions start at under \$200 USD/year, arguably an affordable option for libraries with cataloging volumes between 250 and 7,000 titles a year.

1.2.1 Share Responsibility for Creating Bibliographic Records

OCLC is actively involved in development and strategies that will allow more efficient and automated metadata creation as well as moving toward greater WorldCat inclusion of metadata created by multiple sources. OCLC will proactively pursue the addition of upstream metadata to WorldCat through both of these tactics with the goal of greater title coverage in WorldCat at the time of materials selection and acquisition.

OCLC is also re-engineering existing services such as Cataloging Contract Services to provide more options, price points and greater use of automation to assist libraries in metadata creation for new acquisitions as well as backlog, gift, non-English and non-print and other collections that may be “hidden” due to a shortage of cataloging resources at the library level.

1.2.2.2 LC: For those aspects of operations that extend beyond the Library's immediate mission at the Library of Congress, identify other entities or groups with the interest and ability to assume responsibility for them.

OCLC stands ready to consider any request for help with operations that the Library determines is out of scope within the context of the Library's mission.

1.2.4.1 LC, PCC, and OCLC: Explore ways to increase financial or other incentives for contributions of new bibliographic records and of upgrades or corrections to existing records to the national (and international) share bibliographic and authority databases. (Also 1.3.1.3--authorities)

OCLC is reconsidering its current model for providing incentives for contributing new records, corrections and enhancements to WorldCat and to shared authority files. Once our investigation and analysis of the current model's pros and cons progresses a bit more, we would be glad to be part of joint discussions with LC and PCC.

1.3.2 Increase Re-Use of Assigned Authoritative Headings

OCLC has recently put WorldCat Identities into production. WorldCat Identities, accessible from the “Details” tab of a record in WorldCat.org, is a new end-user service based on authorized headings in WorldCat. There is an Identities page for each person and for many corporate names. Identities is an excellent example of leveraging the many thousands of hours of authority work done by libraries for the use and benefit of information seekers on the Web.

1.3.2.2 LC: Bring together other communities working on problems of author identification; map the issues; and investigate possibilities for cooperation.

OCLC is also working in this area in coordination with the Book Industry Study Group and the ISO working group on the International Standard Name Identifier (ISN).

1.3.3 Internationalize Authority Files

OCLC is an active participant in the VIAF (Virtual International Authority File) project, which now includes the Bibliothèque nationale de France as well as LC and the Deutsche Nationalbibliothek (last month's press release

is available at <http://www.oclc.org/news/releases/200683.htm>). The long-term goal of the VIAF project is to include authoritative names from many libraries into a global service that will be freely available via the Web to users worldwide.

Expose Rare and Unique Materials Held by Libraries

2.1.2 All: Gather and share data on access paths that guide researchers to unique materials as a means to inform best practices for access in a Web environment.

OCLC is interested in partnering with others in this area. Besides WorldCat, OCLC now manages metadata for CAMIO, Hand Press Book, SCIPPIO, ArchiveGrid and a number of CONTENTdm collections, and we would like to learn more about how researchers discover and connect to unique materials on the Web.

2.2 Streamline Cataloging for Rare and Unique Materials, Emphasizing Greater Coverage and Access to a Greater Number of Items

2.2.2 All: Establish cataloging practices that are practicable and flexible, and that reflect the needs of users and the reality of limited resources

And

2.2.4 All: Consider different levels of cataloging and processing for all types of rare and unique materials, depending on institutional priorities and importance and potential use of materials, while still following national standards and practices.

In the context of its Contract Cataloging Services, OCLC will soon debut a “basic black” cataloging service that will provide streamlined, efficient and affordable access-level record creation for these types of materials.

2.5.2 All: Encourage libraries and archives to submit records for rare and unique materials to shared databases such as OCLC.

Libraries that use OCLC’s digital collection management software, CONTENTdm, can add their digital library metadata to WorldCat and share their collections with other libraries and on the Web through Open WorldCat. Longer term plans include integrating metadata for a variety of special collections in WorldCat.

World Wide Web Is the Platform; Position Technology

Recognizing that data stored in separate library catalogs is often not disclosed on the Web, OCLC has introduced Open WorldCat and WorldCat.org to aggregate and more effectively expose library collections on the open Web and drive end-user searches to local library catalogs. An important benefit of WorldCat.org and the Open WorldCat program is making library collections visible outside the boundaries of local catalogs, in effect making local catalogs invaluable components in an interlocking array of discovery tools.

In an environment in which the Web is the platform, the importance of language-neutral identification schemes cannot be overemphasized. OCLC is working on a number of fronts (xISBN, xIdentifiers, etc.) to explore how to use identifiers to relate information objects to one another (using FRBR concepts) for the purpose of improving discovery, navigation of result sets, and delivery for end users. We welcome collaboration with any agency that assigns standard numbers to monographic materials. OCLC is presently engaged in a broad review of identifiers that will assist in embedding persistent, canonical identifiers for library assets in the Web.

OCLC offers a growing number of Web services—called Grid Services—to facilitate machine-to-machine interactions with bibliographic control data. Grid Services, which has a developers’ network to advise it, has three service areas: network services (e.g., search services, metadata extraction), registries (e.g., the WorldCat Registry), and reusable components for developers.

Position Our Community for the Future

4.1. Design for Today's and Tomorrow's User

In the context of the WorldCat Local pilot, OCLC and its partners have created a new kind of “glocalized” catalog with enhanced search and display capabilities like faceted browse, FRBR-ized [Functional Requirements for Bibliographic Records] results sets, and enhanced content like user-contributed reviews and lists, as well as article metadata. Delivery to users is improved by bringing together fulfillment options—electronic, locally held, consortial, and remote—into one interface. The presence of article metadata and digital image collections breaks down former discovery silos. OCLC is currently responding to feedback from pilot partners and also welcomes comments from the library community to help us evaluate how effectively WorldCat Local can encourage user engagement, meet end-user information needs, and position libraries for the future.

OCLC WorldCat Collection Analysis is a tool for supporting evidence-based decision-making in individual libraries and consortia. In addition to using bibliographic data, Collection Analysis makes use of holdings and circulation information. OCLC would be interested in working with partners to move recommendation 4.1.3 and any other efforts to advance a culture of assessment in the library community.

4.2.1 Develop Test Plan for FRBR

OCLC has played a leading role in producing an algorithm that creates FRBR work clusters from WorldCat bibliographic records. We have implemented FRBR work-level search and views into WorldCat.org and WorldCat Local. OCLC has made its work-cluster algorithm freely available, as described at the download site (<http://www.oclc.org/research/projects/frbr/algorithm.htm>). However, the best work clustering results can be achieved with the biggest possible data set, at the network level, that is, for example, WorldCat. The FRBR research and implementation work done at OCLC suggests that if works are clustered, then it is possible from any one record in the cluster to navigate easily to others inside and outside a library collection. Clustered records can also provide the potential of migrating bibliographic information such as subject headings and class numbers into new bibliographic records as new manifestations are published. Our research also suggests that it is possible to achieve FRBRization at the work level without recataloging.

Libraries and organizations with their holdings in WorldCat can benefit from clustering data supplied by OCLC. Their systems then need only a mechanism for ingesting the data and employing it in displays. If desirable, OCLC might serve as the agency for creating and maintaining work-level data, which could be distributed as work table data or records. Pending sufficient interest, an OCLC work cluster service might feasibly be positioned as an OCLC Grid Service supporting a variety of catalogs.

We believe that more progress has been made testing FRBR than the WG may realize, and OCLC will be pleased to partner with others in moving recommendations 4.2.1.1 through 4.2.1.4 forward.

4.3 Optimize LCSH for Use and Reuse

OCLC research into LCSH optimization echoes the WG's assertions with respect to problems with pre-coordinated subject strings. Because LCSH guidelines for pre-coordinating subject strings are complex, the cataloger error rate is high, and developing automated approaches to correction and maintenance of LCSH strings in a large database such as WorldCat is extremely difficult and costly. OCLC supports the recommendations to transform LCSH (4.3.1).

4.3.1.4 LC: Increase explicit correlation and referencing between LCSH terms and LCC and Dewey Decimal Classification (DDC).

And

4.3.3.4 LC and OCLC: Explore ways of reducing creation costs and improving effectiveness by working more closely between DDC, LCSH, and LCC.

OCLC welcomes the opportunity to cooperate with LC on these two recommendations. WebDewey already provides many high-quality intellectual mappings from LCSH to DDC, and selected high-quality intellectual mappings from MeSH to DDC.

4.3.2.1 LC: Work with OCLC and/or other appropriate partners to identify a scheme or product that could take advantage of the power of LCSH and serve as a base to take advantage of terminologies that function in a more accessible environment with broader audiences. The FAST is an example of such an attempt.

Staff at LC, the OCLC Office of Research, and the University of Kentucky have been collaborating on FAST (Faceted Application of Subject Terminology) for some years now. OCLC has built a FAST authority file and an interface to it is available. The research suggests that FAST implementation would provide for easier, more accurate and intuitive application of LCSH terminology and save catalogers time.

4.3.3 Encourage Application of, and Cross-Referencing with, Other Controlled Subject Vocabularies

OCLC has implemented a Terminologies Service to make multiple vocabularies available in a single interface. Terminologies is available to Connexion users or it may be used with a variety of metadata editors. Current terminologies are listed at <http://www.oclc.org/terminologies/>.

In addition, the OCLC Office of Research has invested considerable effort in making vocabularies cross-searchable and interoperable; this initiative is described at <http://www.oclc.org/research/projects/termservices/>. High-quality mappings have been made for DDC, LCC, the ERIC thesaurus, Genre Terms for Fiction, LCSH, LC Children's Headings, MeSH, and the National Library of Medicine Classification. We believe OCLC's intellectual mappings (using human intelligence) are superior to those created through straight statistical correlation techniques.

Third, it may be worth noting that OCLC research and translation partners are also actively involved in mapping the DDC to other vocabularies—examples include SWD (the German subject authority file), Soggettario (the Italian subject authority file), and HILT (High-Level Thesaurus Project), in which a DDC-based interoperability service is used to support cross-searching of heterogeneously indexed collections). OCLC Programs and Research has also explored creating mappings to French subject headings using the LCSH-DDC mappings.

Strengthen the Library Profession

5.1 Build an Evidence Base

The OCLC membership reports--the 2003 Environmental Scan: Perceptions of Libraries and Information Resources (2005); and the recently released Sharing, Privacy and Trust in Our Networked World—all present findings suggesting gaps between end-user expectations of libraries and librarians' understanding of user needs and behaviors. This gap applies as well to the community's understanding of the costs, benefits and value of bibliographic control to end users.

5.1.2.3 All: Work to develop a stronger culture of evaluation and build a cumulative research agenda and evidence base.

OCLC would welcome the chance to help support symposia or continuing education programs in support of this recommendation. As appropriate, some offerings might be made in connection with WebJunction training programs. WebJunction is OCLC's online community for library staff. As previously mentioned, the WorldCat Collection Analysis tool supports evidence-based decisions about collection development and collaboration across libraries.

Again, many thanks for the opportunity to respond to the WG's draft report and recommendations.

Respectfully submitted,

A handwritten signature in cursive script that reads "Karen Calhoun".

Karen Calhoun
Vice President, WorldCat and Metadata Services