

***On the Record Report* Recommendations the Library of
Congress Should Pursue
Over the Next Four Years**

Report to the Associate Librarian for Library Services

September 15, 2009

***OTR Report* Implementation Working Group
Regina Reynolds and Bruce Knarr, co-chairs**

On the Record Report Recommendations the Library of Congress Should Pursue Over the Next Four Years

Contents:

- I. Executive Summary**
- II. Projects**
- III. Contingent Projects**
- IV. Investigations**

- V. Appendices**
 - A. Pilot Project to Use ONIX Data to Create CIP Descriptions
 - B. Sample List of Data for a Crosswalk Registry
 - C. Junior Fellows Intern Project to Create a Database to Describe Special Collections
 - D. Proposal for a Pilot to Incorporate AMS Controlled Subject Terms into LC Catalog Records
 - E. Proposal for a Pilot to Incorporate Externally-Generated Data from LibraryThing into Library of Congress Catalog Records
 - F. Rationale for a Policy to Support Inclusion of Evaluative Data in Bibliographic Records
 - G. Activities to Monitor

I. Executive Summary

Introduction

This report distills the *On the Record Report* (OTR) recommendations into activities that the Library of Congress (LC) should include in its planning process over the next four years. The activities we identified include a mix of new initiatives (some dependent on completion of other activities) and areas warranting further investigation. We also included an appendix listing activities in progress that should be monitored. The report is organized as follows:

- **Projects:** pilots, proposals and projects underway.
- **Contingent projects:** projects, or activities that cannot be started until completion of an in-progress or planned activity.
- **Investigations:** areas where further research and analysis are needed to determine if LC should take further action via projects, partnerships, or other means.
- **Appendices:**
 - **Appendices A – F: Project proposals or descriptions**
 - **Appendix G: Activities to Monitor**, i.e., activities that are in progress or planned (either by LC or groups outside LC). These activities require institutional awareness because they are key activities either directly related to, or in the spirit of OTR.

Policy decisions are needed in at least two areas in order for some of our recommendations to go forward. A policy decision is needed about inclusion of evaluative data in LC bibliographic records. We discuss this need further in Appendix F. Policy regarding the allocation and deployment professional staff resources is also needed, especially in light of the potential disruptions to staff duties that could result from efficiencies listed below.

We recognize the tension between efficiencies and enhancements in the recommendations. Efficiencies include use of automation, partnerships, elimination of redundant efforts, and ceasing some activities altogether. Enhancements include initiatives to bring uncontrolled resources under some kind of control, addition of new data to existing bibliographic records (e.g., reviews, user-generated content, additional vocabularies), and provision of new, additional, or different kinds of data to the information community. While an underlying premise of OTR is that the resources freed by efficiencies can be redirected towards enhancements, it is tempting—as our group has done—to recommend many more enhancements than efficiencies. Enhancements are “sexy;” efficiencies are mundane.

Given the premise that LC will not experience an overall growth in staff resources, the only way enhancements can happen is through the vigorous pursuit of efficiencies, while at the same time preserving the unique contributions to bibliographic control that LC is in the best position to continue to provide. Also in the area of staff resources, we note with both amusement and alarm that a common refrain in our discussions of potential projects was, “All roads lead to our cataloging automation specialist.” In order to increase the

number and pace of Acquisitions and Bibliographic Access Directorate (ABA) automation efforts, it will be important to identify and/or develop additional ABA staff to provide the kind of automation support that ABA now relies almost exclusively on one specialist to provide.

Despite our efforts to distill, collapse, and eliminate, we have likely listed more activities than can be supported due to time and resource constraints. The bulleted items below are those activities we identified as having the highest priority, but the ultimate selection of activities and allocation of priorities and resources should rest with the Associate Librarian for Library Services (ALLS) and the ABA Management Team (ABAMT). It is crucial, however, that each selected activity (whether an initiative, investigation, or monitoring activity) should have an appointed stakeholder, accountable to ALLS.

Highest Priority Activities

- ONIX pilots: only by conducting pilots can LC determine if use of ONIX data is feasible and provides efficiencies. Assess the pilot results to also determine if there are opportunities for batch creation of catalog records for U.S. commercial imprints. Only by exploring efficiencies will LC know if it can afford enhancements. (See Appendix A).
- Once the R2 study on the bibliographic landscape has been completed in September 2009, use the study as another mechanism for determining potential efficiencies, including activities that might no longer be necessary for LC to undertake.
- Appoint “stakeholder” owners for all selected activities: this will help ensure that activities stay on track. We noted that in the LC response to OTR several activities were projected but never accomplished, e.g., the Cataloging Policy and Standards Office (CPSO) project to use software to suggest subject headings (LC OTR Response 4.3.4.1).
- Determine a policy on inclusion of evaluative data in LC records. Any pilots to include evaluative user-generated data cannot go forward without such policy. (See Appendix F).
- Conduct a broad investigation of potential sources of metadata to include sources from within and outside the library community. Consider purchase, lease, re-purposing, partnering with sources, etc. Using or building on existing metadata rather than creating it from scratch is an important potential efficiency.
- Use id.loc.gov as a model for additional efforts that make LC data freely available to others to use creatively and report to the community. This model allows any individual, group, or organization in the broad community to bring the full force of its creativity to bear on LC data, potentially leading to new uses and tools.
- Establish a mechanism to solicit funds to use as grants to commission research that will benefit LC and its community. An example is given in the Projects section

where this approach is suggested as a means to obtain research about a new metadata carrier.

Next Steps

The *On the Record Report* and the efforts that followed at the Library to consider its recommendations have been invaluable in focusing attention on the choices needed to move the Library forward. We intend the recommendations offered in this analysis to help transition attention from the work of considering the report to taking appropriate concrete actions. Further work implementing these recommendations will best be accomplished by ALLS and ABAMT selecting and prioritizing projects and activities to pursue; appointing stakeholder owners for them; and linking the activities to the performance budget.

OTR Report Implementation Working Group

Co-chairs:

Regina Reynolds and Bruce Knarr

Members:

Victoria T. Behrens

Colleen R. Cahill

Kurt Carroll

Daniel Chudnov

Jennifer Harbster

Phil Michel

Jennifer Baum Sevec

Michelle Springer

II. Projects

This category includes projects, pilots, and activities.

OTR Report 1: Increase the Efficiency of Bibliographic Production and Maintenance

CIP and/or ONIX:

1. Update LC's ONIX to MARC crosswalk. (Pilot crosswalk completed, May 2009)
2. Conduct pilot projects to use ONIX data already being supplied to LC by publishers. John Wiley and Cambridge University Press have been selected for the quality and quantity of their CIP records. Projects will be conducted at both LC and at the National Library of Medicine (NLM). (LC pilot scheduled for June 8-30, 2009)
3. Create Resource Description and Access (RDA) records using ONIX data. During RDA testing by LC and NLM, staff members are to analyze the records to determine if they meet RDA's goal of allowing more flexibility with publisher- and vendor-supplied data. (2010)
4. Complete the CIP workflow analysis (already begun in the U.S. and Publisher Liaison Division (USPL)) to determine optimum data flows and use of automation for the CIP program. (In progress, 2009).
5. Depending on the outcome of the ONIX pilots, investigate the feasibility of using the ONIX data some publishers send to *Books in Print* to obtain ISBN as a source for CIP metadata. (2009-2010).
6. Evaluate the results of the above testing and analysis to assess the feasibility of a fully-automated CIP process and if, feasible, develop a pilot to test automation of CIP. (2010)

Other Projects

1. Establish a Web page registry/clearinghouse for existing Crosswalks or expand and update the information provided by the Network Development and MARC Standards Office (NDMSO). See Appendix F for a list of items for possible inclusion. (2009)
2. Create and follow a timetable for adapting existing LC and Program for Cooperative Cataloging (PCC) courses into distance learning courses. Make adapting authority classes the first priority. ABA's training division has already experimented with Elluminate, Blackboard, and similar tools. Also responds to OTR Report #5 Strengthen the Library and Information Science Profession. (2009-2010)
3. Develop a new component for the Automated ISSN Register program so that the program can accept online ISSN applications and convert the publisher-supplied metadata to preliminary MARC serial records. (2009-2010)

OTR Report 2: Enhance Access to Rare, Unique, and Other Special Hidden

Materials

1. Create a searchable inventory database to hold information about all special collections at the Library of Congress, including but not limited to information gathered during the survey that formed part of the strategic planning process. Utilize the assistance of two Junior Fellows recruited for this project. Make the database available via the Internet. Use the database to determine collections needing conversion tools and to set priorities for digitizing and creating metadata. See Appendix C. (In progress, June–August 2009)
2. Create tools to facilitate efficient creation, conversion, and discovery of finding aids, using Encoded Archival Description (EAD) for materials as appropriate. (2009-2011)
3. Test and evaluate DCRM(B) and DACS workflows. Determine the value and decide on adoption. (2010-2011)

OTR Report 3: Position our Technology for the Future

1. Offer from five to ten grants (from funds solicited from interested parties) to support research projects to develop a more flexible, extensible metadata carrier. Grant proposals should be invited from a broad range of academic, research, and vendor communities. Recipients must produce test data sets and share them with LC and the community. (2010-2012)
2. Commission a study or enlist a library or information science graduate student to perform research to identify any new descriptive practices that are needed to support emerging uses of bibliographic data such as those seen in new discovery environments as well as determine whether there are any cataloging rules or practices that impede discovery in new environments. A less systematic and more informal way to obtain data is to post this question on blogs or use other social networking tools to elicit information from those who are working with existing bibliographic data in new discovery environments. In the study requirements, indicate that the results of the University of Wyoming Libraries survey on “Effects of a Next-Generation Catalog on Catalogers” and any similar surveys or articles should be examined to determine topics that might be included in the research study. (2010)

OTR Report 4: Position our Community for the Future

1. Implement the pilot project to incorporate American Mathematical Society (AMS) Subjects/Thesauri in CIP records for AMS publications. It is estimated that forty titles per month will be enriched in this way. Assess the AMS pilot and if it has provided subject access points supplemental to LSCH at minimal cost, identify and carry out similar pilots with vocabularies other organizations/publishers such as IEEE, AHA, and other as appropriate. See Appendix B. (2009)
2. Conduct a pilot project to test means to incorporate user-generated data into Library of Congress catalog records; specifically, the incorporation of character names derived from LibraryThing (LT). See Appendix D. (2009-2010)
3. Pilot digital table of contents (DTC) for Japanese and Korean works under a project began with BEAT and through the direction of a section head in the Asian and

Middle Eastern Division (ASME) of ABA. The pilot may be further expanded for Arabic works with support from ASME. (2009-2010)

4. Develop a policy to support inclusion of evaluative data in bibliographic records. Such a policy is a necessary precursor to any project that would incorporate evaluative data into records. See Appendix F. (2009-2010)

OTR Report 5: Strengthen the Library and Information Science Profession

1. Distill from annual reports, Congressional testimony, and other statistical information, as well as from the Library's mission statement, a statement of the value of LC's services, in monetary, educational, and cultural terms. Publicize the statement on the Library's Web site and in other appropriate venues. (2010)
2. Communicate to ALA, LIS schools, and other professional groups and institutions LC's interest in and support for their taking the lead in strengthening the library and information science profession. (2010)
3. Recommend that in addition to making Catalogers Learning Workshop (CLW) course materials such as the Basic Subject Cataloging Using LCSH Cataloging Skills (CCT) course (<http://www.loc.gov/catworkshop/courses/basicsubject/index.html>) available on the LC Web site, we also make this material (slide shows, instructors manuals, etc.) available on the forthcoming Library iTunes U page, which will feature content and educational materials from the Library's collections. This will increase awareness of the material, allow people to discover the material as they look for other educational materials on the iTunes U site, and allow people to subscribe to the page that would alert them when new materials are uploaded or revised. (Library iTunes U page launched on June 30, 2009)

III. Contingent Projects

These projects are contingent on an in-progress or planned activity.

OTR Report 1: Increase the Efficiency of Bibliographic Production and Maintenance

1. Investigate the potential conversion of data submitted for Copyright as the basis for MARC 21 bibliographic records when Copyright processing has become stabilized. (2011-2012)
2. Analyze the results of the R2 bibliographic landscape study and use this information to develop recommendations for change. (2009-2010)
3. Review cataloging production, overlaps, and efficiencies at LC in conjunction with the first-year assessment of the ABA reorganization. (2010)

OTR Report 2: Enhance Access to Rare, Unique, and Other Special Hidden Materials

No contingent projects defined at this time.

OTR Report 3: Enhance Access to Rare, Unique, and Other Special Hidden Materials

1. Study the use of identifiers for RDA elements as part of the RDA testing. (2010-2011)

OTR Report 4: Position our Community for the Future

1. Investigate what kinds of information could be leveraged from the Automated Call Slip (ACS) Project once it is fully implemented. For example, examine the feasibility of suggesting other titles to users or highlighting works on popular topics once sufficient data, appropriately masked regarding personal information, are accumulated. (2010-2011?)
2. Seek input from systems developers or experimenters who worked with RDA records created during the test period in actual or experimental environments that use Functional Requirements for Bibliographic Records (FRBR) displays following the completion of the RDA testing conducted by the U.S. national libraries. Investigate topics such as whether FRBR displays work for all formats and types of resources and changes that might be suggested for how FRBR is applied. (2011-2012)

OTR Report 5: Strengthen the Library and Information Science Profession

1. Analyze the report of ALCTS Technical Services Directors of Large Research Libraries Discussion Group on the development of key measures of the costs, benefits, and value of bibliographic control and the interdependencies of various groups of participants in the bibliographic control area. Following this analysis and analysis of the LC-commissioned R2 report, determine what further actions LC should take. Consider using these results to determine a direction and focus in order to charge a working group “to work together ... to develop a shared frame of reference and common design goals for a coordinated renovation of the shared bibliographic apparatus.” (2009-2010)

IV. Investigations

These are areas for further investigation and analysis. Projects could be developed following investigations with positive outcomes.

OTR Report 1: Increase the Efficiency of Bibliographic Production and Maintenance

1. Charge all ABA Divisions to evaluate any existing sources of external metadata currently in use and investigate new sources of metadata, including abstracting and indexing services, IMDb (Internet Movie Database), and similar sources in their areas of responsibility. Ask ABAMT to determine the most promising sources and conduct pilot projects using these sources. Evaluate the pilots and use the results of the evaluations to develop a set of best practices about finding and using external metadata sources in a systematic and directorate-wide manner. (2010-2011)
2. Consider the grant-funded approach suggested in Section 3.1 as a means to support studies to establish empirically the correlation between user behavior and the content of bibliographic records. OCLC has done some work in this area but its work is more general than what is needed to identify little-used or poorly understood elements of catalog records. Other options would be to suggest library school research projects or to use an LC contract to obtain these data. Some data may possibly come out of RDA testing studies. (2010-2012)
3. Continue to explore new tools and techniques for sharing bibliographic data and Library content through the Web using new standards and approaches (e.g., Linked Data, RDF and SKOS, OAI-PMH and OAI-ORE, SRU, Atom, OpenSearch) as they develop. Engage both NDMSO and the Office of Strategic Initiatives (OSI) in leading discussions on this topic within Library Services. (2010)

OTR Report 2: Enhance Access to Rare, Unique, and Other Special Hidden Materials

1. Explore using Geospatial Information Systems (GIS) for searching and displaying holdings of all types of materials. (2009-2011)
2. Research and encourage use of text processing tools. Study Automatic Metadata Generation Applications (AMeGA) recommendations; consider investigating National Digital Information Infrastructure and Preservation Program (NDIIPP) or eDeposit content. (2009-2010)
3. Evaluate data and content visualization tools, e.g., Many Eyes, Seadragon (used in the World Digital Library (WDL)). (2009-2010)
4. Investigate use of collections in union/shared catalogs, search engines, and online registries. (2009-2011)

5. Pursue more federal and state government cooperation, such as was started by NDIIPP; content came to the Library but only for preservation; determine processing and access options. (2010-2011)
6. Investigate practices and policies related to cataloging levels in Special Materials collection/cataloging area. Convene a meeting of internal LC stakeholders to share information. (2009-2010)
7. Consider guided search tools/recommender system to assist researchers to locate materials in various locations. (2009-2012)

OTR Report 3: Position our Technology for the Future

1. Explore the feasibility of adding geographic coordinates found in Geo-Names to Library of Congress Subject Headings (LCSH) place names in bibliographic records. If feasible, do a pilot project to test this concept further. (2010)

OTR Report 4: Position our Community for the Future

1. Further expose and interconnect vocabulary and thesaurus data via URLs by using the example of LCSH terms being linked to terms in RAMEAU (a French subject heading vocabulary closely coordinated with LCSH) by linking to other appropriate vocabularies. (2010-11)
2. Implement the project to test software that would suggest subject access points and call numbers for digital texts. LC's response to OTR states, "[the] CPSO report on pre- and post-coordination of LCSH...estimated that such a capability could be tested by the end of fiscal year 2008..." For example, investigate such automated (or semi-automated) indexing products as Teragram Categorizer and Teragram Taxonomy Manager <http://www.teragram.com/solutions/categorizer.htm> or NLM's Medical Text Indexer <http://ii.nlm.nih.gov/mti.shtml>. Consider pilots in specialized areas, e.g., physical sciences, or other disciplines with specialized vocabularies. (2009-2010)
3. Investigate the pros, cons, and feasibility of moving as many of LC's Web-based table-of-contents (TOC) data into catalog records as possible. Increase the incorporation of table-of-contents data into LC catalog records, especially by including the TOC data in the record instead of supplying an outside link to the data. This project is based on LC users' experience and the OCLC report on *Online Catalogs: What Users and Librarians Want*. The OCLC report states, "End users rely on and expect enhanced content including summaries/abstracts and tables of contents." (2010)

V. Appendices

Appendix A: Pilot to Use ONIX Data for Descriptive Portion of CIP Records

Text provided by the developer of the current ONIX to MARC crosswalk, and communicated to ABA staff June 3, 2009.

Project status: In process. Pilot project duration: June 8, 2009 – June 30, 2009. Statistical and subjective data will be recorded by the pilot participants.

In the Report of The Library of Congress Working Group on the Future of Bibliographic Control, *On the Record*, section 1.1, Eliminate Redundancies, has several recommendations for using externally available bibliographic data and for further automating the CIP process. With this in mind, the Acquisitions and Bibliographic Access Directorate (ABA) will conduct a pilot project in June 2009 designed to make available ONIX data being received from publishers to the Electronic Cataloging in Publication (E-CIP) program. A test involving ONIX data from two publishers, Cambridge University Press and Wiley, will test several aspects:

- The availability of ONIX data for items in the CIP stream
- The usefulness of the data in cataloging
- Any problems or unexpected results from converting the data from ONIX to MARC
- Changes that would be needed to the CIP workflow
- Additional information that can be extracted from the ONIX data that would not normally be provided in MARC records

A virtual test section will be established in the E-CIP Traffic Manager and incoming CIP applications from the two publishers will be diverted to this virtual section (except for items for the National Library of Medicine and the National Agricultural Library) for descriptive cataloging processing. If an ONIX record is found (based on matching the ISBN of the forthcoming book with ISBNs in the ONIX data), the data will be converted immediately and a MARC record created. From here, the catalogers involved will compare the resulting record to the publisher-supplied information from the electronic galley to look for differences or any missing/incorrect elements. Should there not be an ONIX record for the forthcoming book, the CIP application will be forwarded to its original destination for normal processing.

In addition to the basic bibliographic record, the table of contents, if provided in the ONIX record, will be provided in the 505 field of the MARC record with first indicator value '8' (no display constant) and the legend "Machine generated contents note:." No attempt will be made to convert the field into a "perfect" contents note, so elements not normally found in a regular AACR2 contents note will be in the 505, such as the words "Chapter," "Part," etc. as well as sections of the table of contents like the introduction, bibliography, etc. that are not normally mentioned.

Additionally, ONIX records frequently contain summaries (called descriptions in ONIX) and also on the E-CIP application there is a space for the publisher to provide summary

information. If either of these is present, the conversion program will include them in separate 520 fields, quoted to indicate that LC did not create the summary and with "-- Provided by publisher." at the end to indicate the source. The catalogers involved will read any summaries and if they fit guidelines for including summaries in E-CIP records, the summaries may be left in the record. There is a potential for duplicate summaries to be created. If the summaries are very similar or identical, the cataloger would delete one of them.

Once the rest of the descriptive cataloging process is completed, the catalogers involved will create a report on the results, provide any impressions they wish to relate, and will then forward the E-CIP to the original destination for completion of subject cataloging and end-stage processing. It is anticipated that this diversion and conversion will not add significant processing time to the E-CIP.

As the pilot proceeds, any needed changes in the conversion application will be made as will any needed changes to the pilot workflow. Reports will be gathered weekly so that the pilot committee can be kept up to date on all of the issues.

Appendix B: Examples of Potential Data for a Crosswalk Registry

Project Status: Proposal

- Getty
 - CDWA to CCO, VRA, USMARC, DC, Object ID, FDA Guide, CIMI, and EAD; DC to EAD; USMARC to EAD; ISAD(G) to EAD; EAD to ISAD(G)
 - http://www.getty.edu/research/conducting_research/standards/intrometadata/crosswalks.html
- CHIN Humanities Data Dictionary
 - CHIN has mapped the Humanities Data Dictionary to other existing standards. This crosswalk is based on the Getty Standards Program's *Crosswalk of Metadata Element Sets for Art, Architecture, and Cultural Heritage Information and Online Resources*. It maps among:
 - CHIN Humanities Data Dictionary ,
 - *Categories for the Description of Works of Art*
 - *Object ID*
 - *CIMI Access Points*
 - *VRA Core Categories*
 - Dublin Core
 - USMARC
 - <http://www.chin.gc.ca/English/Standards/appendix2.xls>
- IEEE (from 2002)
 - DC to LOM
 - http://ltsc.ieee.org/wg12/files/LOM_1484_12_1_v1_Final_Draft.pdf
- OCLC (2002 -2008)
 - CanCore to SCORM; ONIX to MARC 21; LOM to DC; GEM to MARC
 - http://www.oclc.org/research/projects/mswitch/1_crosswalks.htm
 - Various efforts to crosswalk from an R&D perspective
 - <http://www.oclc.org/research/researchworks/default.htm>
 - ONIX to MARC (66 slides – pilot: Jan.-June 2008)
 - http://www.oclc.org/us/en/multimedia/2009/files/ONIX_to_MARC.ppt
 - ONIX to MARC (33 slides)
 - http://www.wils.wisc.edu/events/peer08/WiLS_Next_Gen.ppt
- LC
 - EAD Crosswalks: ISAD(G) to EAD; EAD to ISAD; DC to EAD; USMARC to EAD
 - http://www.loc.gov/ead/tglib/appendix_a.html
 - MARC/DC and DC/MARC; DC/MARC/GILS
- NC State U.
 - Petite Schema crosswalk (broken link)

www.lib.ncsu.edu/cataloging/metadata/PetiteCrosswalk.html

Appendix C: Pilot to Incorporate AMS Controlled Subject Terms into LC Catalog Records

Project status: Approved, awaiting preparations and implementation

Relevant Sections within OTR:

- 4.1.2 Integrate User-Contributed Data into Library Catalogs
- 4.3.3 Encourage Application of, and Cross-Referencing with Other Controlled Subject Vocabularies

Goal: To facilitate bibliographic access and enrich bibliographic records by adding society/publisher controlled subject terms provided in the CIP galley. The terms are identified by the society/publisher in coordination with the author(s)/editor(s) and will serve as searchable subject terms in addition to the LCSH already provided by LC catalogers. These data are likely to be of particular interest to specialists in the field, and may provide alternate access terms in addition to LCSH provided in LC catalog records.

Proposal: Pilot a project using American Mathematical Society (AMS) controlled subject terms provided in the CIP galley.

Phase 1: AMS staff will supply their controlled subject terms in the front matter of the galley. LC catalogers will copy the terms into 650 fields on CIP record. Indicators for the AMS terms will be coded zero for indicator 1 and a publisher specific number for indicator 2, per the Policy and Standards Division (PSD) and NDMSO (Network Development and MARC Standards Office). The pilot is proposed for the Science, Technology, and Medicine (STM) section which is the predominant recipient of AMS galleys for CIP. The proposal is to start with AMS galleys received from a date mutually set between AMS and LC.

Phase 2: After the Voyager upgrades have been completed (late in 2009), cataloging automation specialist will augment Traffic Manager (ECIP) to allow AMS staff to input their controlled subject terms into fields in the Traffic Manager Web form. The AMS terms will then automatically import from Traffic Manager into the LC bibliographic record.

Phase 3: Evaluate the pilot. STM catalogers will catalog AMS records and incorporate the AMS terms into the bibliographic record as proposed. ATM will report the number of AMS galleys received and completed. Catalogers will assess the quality of the records. If the pilot is deemed a success, the pilot will be expanded to incorporate other controlled vocabularies from societies/publishers who participate in the CIP program.

Phase 4: Scale the pilot to incorporate other controlled vocabularies from identified societies/publishers who participate in the CIP program (i.e., IEEE, AHA, etc.). These data are likely to be of particular interest to researchers, scholars, and academic librarians. The data provide an opportunity to capture descriptive subject terms as determined by the publisher in consultation with the author(s). The society/publisher provided controlled subject terms are not currently captured in LC catalog records, and provide an opportunity for alternative access points from LCSH. The data are not likely to conflict with other LC data.

Potential level of effort:

- Technical: Likely to be low. Phases 1 & 2 require minimal input from AMS staff and LC catalogers that is similar to current input. The proposed workflow saves LC catalogers the added steps of looking up the AMS subject codes online. The cataloging automation specialist will need time to enhance Traffic Manager to automate the process. His initial assessment is that minimal effort will be required. Phases 3 & 4 require some evaluation, but none that is particularly time consuming. STM currently evaluates their records through peer review. Catalogers will be asked to evaluate the records that they work on. The number of records received and cataloged can be easily queried in the Voyager cataloging module through existing search functions.
- Policy: Likely to be moderate. PSD and NDMSO have been consulted from the onset.

Steps/Issues:

- Explore further with specialist the preferred formatting of publisher data to submit controlled vocabularies to CIP.
- Discuss time and timeline with specialist for augmenting Traffic Manager to automate adding publisher controlled vocabularies into the bibliographic records via Traffic Manager.
- Arrange for the pilot with STM Section Head.
- Inform appropriate CIP management and staff that the controlled vocabularies will appear in the galleys and should remain in the data submitted for cataloging.
- Arrange for an additional subject specialist cataloger to be added to STM workflow in Traffic Manager.
- Draft instructions for AMS staff as well as the STM Section.
- Agree on a start date with AMS.

Appendix D: Junior Fellows Intern Project to Create a Database to Describe Special Collections

Project Status: In progress

Junior Fellows Intern Project
June 1 – August 7, 2009

Project leaders: Morgan Cundiff, Bruce Knarr, Regina Reynolds

Interns: Brian Norberg (Indiana University), Jennifer Green (University of Kentucky)

Project Goal

Establish a comprehensive database that will eventually describe all Library Services special collections.

A comprehensive database for special collections is an essential tool for support of management decisions, for support of workflow and processing activities, and eventually for support of end user access to information about the collections. The database will be especially useful as the Library seeks to improve access to “hidden collections.”

Background

The rationale for the creation of this database has been articulated in two recent Library planning documents:

- 1) The idea was first proposed in the 2007 Strategic Planning Working Group Discussion Document (Performance Goal 2.A.2) where it was stated: “We therefore propose a database for special collections into which all divisions could enter information about each of their special collections.”
- 2) The Library’s On the Record Coordinating Group proposed that the project be undertaken as part of the summer 2009 Junior Fellows Intern Project (proposal attached).

Deliverables

- 1) Establish a Web-based database that provides information on special collections owned and managed by Library Services. The database will include various data elements including the following: name of collection, description of collection, reading room contact information, whether or not the collection is represented in the ILS, a description of any existing printed or electronic finding aid or database (with location of any files), and whether or not the collection is already digitized or is a candidate for digitization in the future.
- 2) Submit a Project Summary Report to the Associate Librarian.

Methodology

1. Devise a method for identifying and extracting existing ILS records that describe special collections.
2. Interview appropriate staff in the eight divisions that make up the Library Services Special Collections and Services sub-directorate, and also the six divisions from the General Collections and Services sub-directorate (that also holds special collections). Identify non-ILS collection descriptions (published books, inventories, lists, databases, spreadsheets, card catalogs, accession files, etc).
3. Harvest (and reformat as needed) the data gathered in Steps 1 and 2 for import into the special collections database.

Appendix E: Pilot to Incorporate Externally-Generated Data from LibraryThing into Library of Congress Catalog Records

Project Status: Proposal

Goal: To expand records with external data elements not captured by traditional cataloging by libraries or publishers that would add value to the records for users of the collection.

Relevant Sections within OTR:

4.1.1 Link Appropriate External Information with Library Catalogs

4.1.2 Integrate User-Contributed Data into Library Catalogs

Objective and Background:

As a pilot project, we propose to add character names to Library catalog records of works of fiction in English, deriving data from an external source that provides access to this data through an API (application programming interface) at low or (preferably) no cost. These data are likely to be of interest to school and public libraries, are not currently captured in LC catalog records, and are not likely to conflict with other LC data. Character name data may not be the most demanded data element to add to catalogs; however, works of fiction are usually not cataloged by subject. This effort will add more searchable elements to those records, and it is not uncommon for patrons to try to locate works of fiction for which they can only remember plot or character names. This pilot will also demonstrate the value of data provided by readers of a work, and is not likely to generate objections related to the perception of LC endorsement of subjective data (such as links to book reviews) that have been voiced in the past. This data set is a good place to begin to work out the presentation, policy, and technical issues involved with user-generated data from an outside source.

Pilot Parameters:

- Ensure that use of the data set of character name is without restrictions, or that the Library can respect any license on the data, particularly when engaging in record distribution beyond the Library.
- Data should be searchable and/or crawlable.
- Data source would be one that collects user-generated data on works of fiction written in English.
- Data should be in English and therefore less likely to contain non-Roman characters.
- Data should be free or very low cost, preferably available via API.
- Data do not have to be fielded but have to provide a way to sort by type of data to limit set to named characters. For example, a machine-generated list of undifferentiated capitalized phrases which could include names of people and geographic locations is not what is sought for this pilot.
- For the initial pilot, data would be added in a note to the MARC 500 field. Following evaluation by CSPO, a successful outcome could spark a request to MARBI to define a new indicator value for the 511 field (currently the field for cast and performers data) and character name data could be placed there. Data

added to the note field would include information on the source of the data.

Identified Potential Data Source:

- **LibraryThing** (LT) (www.librarything.com) offers an API for their common knowledge project which captures, among other data elements, People/Character (characternames) data. (See the sample XML block of data for the book *Jonathan Strange & Mr. Norrell* <http://www.librarything.com/services/rest/1.0/?method=librarything.ck.getwork&id=1060&apikey=d231aa37c9b4f5d304a60a3d0ad1dad4>.) Note: This is not a proposal using “LibraryThing for Libraries” which is a separate product.
- LT character name data are not fielded. Entries would have to be entered as a string: ex: Jonathan Strange, Gilbert Norrell, the gentleman with thistle-down hair, Stephen Black, Lady Pole, Mr. Childermass, Mrs. Pleasance, Vinculus, The Raven King.

Potential Level of Effort:

- Technical: Likely to be low; LT offers the data via API; LC has techniques for automated matching and record update.
- Policy: likely to be moderate (see below)

Policy/Business Issues to Resolve:

- Should LC link to the data (*a la* TOC information), or import the data directly into the catalog record, or both? If only imported, the data would be “frozen” and not updated in LC’s catalog, but could be updated in LT.
- Licensing: Is there conflict between the licenses of the LT data (“LibraryThing data from our Common Knowledge project is covered by the Common Knowledge License, equivalent to the Creative Commons Attribution-Share Alike 3.0 Unported”) and the conditions/licenses under which LC distributes catalog records?
- If LC records with LT data are distributed, are there issues that will arise when those records begin to involve OCLC?

Next Steps/Issues:

- Explore further with specialists to flesh out a proposal and issues to be resolved. For example: “determine the business rule of maintaining this data”—has ramifications as to whether the data are to be in the catalog as a link to a separate page (*a la* TOC information) or actually imported into the record.
- LC would need to approach LibraryThing and discuss a pilot proposal. Data would have to be attributed to LibraryThing.
- Office of the General Counsel (OGC) would have to review the agreement.
- The issue of licensing would have to be discussed and clarified amongst the parties (see above).
- Determine the parameters of data to be pulled (by ISBN?) to match a bounded set of LC records for fiction. Consider the LC records for a select set of publishers.

(Could be a topic of discussion with LT).

- Determine communications plan on what this pilot is, what its objectives are, etc. to answer likely questions from the catalog record community.

Performance Measures/Discussion Points

- Quantitative: Determine possible measurements of how often data are accessed after imported into the records.
- Qualitative: capture and record process and issues resolved when engaging in this type of project to ease extensibility to other similar data set projects.
- Identify measures to determine if LC record users/subscribers find this type of data augmentation desirable and useful.

Appendix F: Develop a Policy to Support Inclusion of Evaluative Data in Bibliographic Records

Project status: Proposal

If the Library supports the development of systems capable of relating evaluative data, (as recommended in *On the Record*) then it should support the inclusion of this type of data. At present the source of catalog record inclusion of evaluative data (in the form of reviews) occurs in conjunction with copy cataloging. There have also been projects in the past to ingest reviews from the *Handbook of Latin American Studies*, and from various partners via the Bibliographic Enrichment Advisory Team (BEAT) (through ingest and/or linking to reviews).

Before specific pilots can be proposed to incorporate evaluative data, senior management must determine the desirability and acceptance of risk related to evaluative data and supports the development of a *policy framework* under which future initiatives related to evaluative data could successfully operate and be sanctioned. Evaluative data can be negative as well as positive and therefore may be (and has been) perceived as having negative connotations for the work evaluated. In addition, the selection of sources and appearance of such data must be carefully constructed so as not to appear to convey endorsement of a commercial entity or the data in question.

A comprehensive policy on this matter would establish how to effectively mitigate risks (such as using data from established vetted partners), address aspects such as copyright, privacy, removal requests, attribution, disclaimers, and the degree to which the evaluative data can be edited (if at all), and would provide operational guidance for institutional implementations. Special attention will have to be applied to selection and attribution of the sources of such data so as not to give the appearance of endorsement, and this consideration should be an integral component of the development of systems that relate to the data.

Without such a policy, responses to concerns expressed by authors/publishers and requests for takedowns may be applied unevenly. Past experiences related to inclusion of book reviews in the BEAT project illustrates the need for such policies. Once developed, this policy should be communicated both internally and externally. The policy must be owned by a division responsible for its maintenance and evolution.

Appendix G: Activities to Monitor

Monitoring may in some cases lead to investigations or projects. In other cases, monitoring may identify projects in need of resources, policy attention, or further direction. Initiatives to be monitored are not products of the OTR group. Rather, they are activities requiring institutional awareness because they are either directly related to, or in the spirit of the OTR report. Monitoring could lead to proposing a related pilot or project; to ensuring that a project did not founder; or to identifying the need for a policy decision or change in resource allocation.

ALCTS Task Group on the LCWG Report

A formal liaison has been established with this task group and LC should continue to monitor the group's activities and consider collaborations when appropriate. The task group has selected ten of the *On the Record* recommendations for its attention: making bibliographic data available earlier in the supply chain; development of a more flexible extensible metadata carrier; creation and sharing of bibliographic and authority data; improvement of the standards development process; linking appropriate external information with library catalogs; realization of FRBR; optimizing LCSH for use and reuse; bringing key participants together; furtherance of research; and meetings with LIS educators. The task group's wiki is located at: <http://www.alcts.ala.org/implement/>

OTR Report 1: Increase the Efficiency of Bibliographic Production and Maintenance

1. Continue support for LC's participation as a founding member in the development of the Virtual International Authority File (VIAF). Monitor progress with an eye towards leveraging as quickly as possible any developments that could provide workflow efficiencies. (2010-2011)
2. Assess the progress of ISO 27729 (International Standard Name Identifier, draft standard available summer 2009) to determine if its implementation will help meet the need for more sources of existing name authority information. Identify potential sources and partnerships for sharing name authority and author identifications, for example, OCLC's identities project, publishing and authors' rights community efforts such as that undertaken by CISAC. A lengthy list of projects is found at this URL <http://repinf.pbworks.com/Author-identification>. Determine if LC can or should take on any kind of leadership role. (2010-2011)
3. See also item 4. Support RDA testing in OTR 3 below.

OTR Report 2: Enhance Access to Rare, Unique, and Other Special Hidden Materials

1. Monitor recommendations forthcoming from LC's Digital Library Content Group (DCLG) on an implementation plan for digitization that will articulate objectives, prioritization principles, and criteria as they relate to digitization of Library materials.

2. Monitor progress on the digital repository for LC collections, including those in special collection areas. (2009-2010)
3. Monitor implementation of a metasearch engine that will work across the Library's various databases. (2009-2011)
4. Monitor completion of the development of the XML data-store pilot for bridging different types of databases. (2009-2012)
5. Monitor the Geography & Map Division's Multi-sheet Map Series pilot to create a sheet level inventory of 200,000 sheet maps of Africa. (2009-2012)
6. Continue current practices for record submission while exploring new opportunities; watch for RDA and MIC developments and continue participation
7. Sustain standards for EAD, MARC, METS, MODS, MIX, PREMIS, etc. Set example by continuing to offer LC data that encourage new tools/catalog/record sharing developments (LCSH in SKOS).
8. Review the results of the pilot project (June – December 2009) conducted by the Internet Operations Group, in collaboration with the Office of Communications, to improve the quality of topical Wikipedia articles that relate to the unique primary sources in the Library's collection. The focus of the Library's participation will be to add relevant primary source links as cited sources or external links with existing Wikipedia articles. This pilot can provide valuable information about alternative ways to provide access to unique items in the Library's collection.

OTR Report 3: Position our Technology for the Future

1. Study the response to LC's *Authorities and Vocabularies* service. Monitor the stated expansion of id.loc.gov to other vocabularies, especially name authorities and such vocabularies as the *Thesaurus of Graphic Materials*, geographic area, language and relator codes, preservation events, and roles. Consider other sources of data to be added. Monitor community reaction by studying comments posted to id.loc.gov, blogs, and articles in the professional literature. (2009-2010)
2. Follow up on the example of LCCN Permalink, a persistent URL service for creating links to bibliographic records in the Library of Congress Online Catalog using the Library of Congress Control Number (LCCN) by investigating whether other types of standard identifiers can be included in LC bibliographic records. (2010-11)
3. Support proposed project to attach linked references to the Law classification for Native American tribes and explore extending this project to other areas if successful. (2010-2011)

4. Support the testing of RDA by encouraging widespread participation by LC staff in various ways: cataloging records, evaluating the effectiveness of the records, and experimenting with records that have been created. (2010-2011)

OTR Report 4: Position our Community for the Future

1. Follow-up on the Prints and Photographs Online Catalog (PPOC) records which provide URLs that link to Library images hosted in Flickr with the message that additional information may be available there (i.e.: “Additional information about this photograph might be available through the Flickr Commons project at http://www.flickr.com/photos/library_of_congress/2179129718”) by exploring other external sites that might provide opportunities for similar projects.

OTR Report 5: Strengthen the Library and Information Science Profession

1. Ensure that to the extent possible, presentations on bibliographic control topics at LC are taped for Webcasting and podcasting.

Continue to make LC videos on topics of interest to the library and information science profession available not only on LC’s Web site but also on sites such as YouTube, iTunes U and similar sites.