

MLA Liaison Report to BCC from the ALA Annual Meeting New Orleans, LA, June 23-28, 2011

LITA/ALCTS—CCS Authority Control Interest Group (ACIG)
Sunday, June 26, 2011, 1:30-5:30PM

Open Meeting

The theme of the open meeting was *Authority Control in the Next Generation* and consisted of three presentations, plus a report from the Library of Congress.

1. LC Update to the Authority Control Interest Group – by Janis L. Young, Policy and Standards Division (PSD), Library of Congress (LC)

RDA: The U.S. National Libraries will implement RDA no earlier than January 1, 2013. This date is contingent upon satisfactory progress toward the goals identified in the report of testing community. That being said, some LC catalogers who were involved with the test will start using RDA for new cataloging in November 2011. Now that a decision has been reached on implementation, PSD and PCC will resume work on the Library of Congress Policy Statements (LCPS). These are available on the RDA Test Web site (http://www.loc.gov/catdir/cpso/RDAtest/rda_lcps.html), the RDA Toolkit (no subscription required to view LCPS, <http://access.rdatoolkit.org/>), and through Cataloger's Desktop (subscription required, <http://desktop.loc.gov>). In response to the recommendation that MARC21 be replaced, Deanna Marcum, Associate Librarian for Library Services at LC, has created the Bibliographic Framework Transition Initiative. For more information, see <http://www.loc.gov/marc/transition/>.

Romanization: New Romanization tables are available for Vai and Judeo Arabic. The Thai romanization tables are currently being reviewed. Some older tables that were in WordPerfect format will be converted to Microsoft Word and will soon be available on the PSD Web site.

Virtual International Authority File (VIAF): There are now 18 participating institutions in VIAF, with another 4 institutions in the testing stages. VIAF now contains name and corporate authority headings and is being hosted by OCLC.

LC Authorities (<http://authorities.loc.gov>): Labeled displays in English (as opposed to MARC21 fields) are now available at the LC Authorities Web site. RDA elements are also able to be displayed and if an authority record was created using RDA, this will also be indicated. The Web site also indicates if a heading is appropriate for use in the LC Genre/Form Thesaurus (LCGFT). Former, but now invalid, LCCNs are also displayed on LCGFT records.

PSD Staff and Workflow: Senior Subject Specialist Paul Weiss retired in February 2011 after 37 years of service. There are now only 3 subject specialists in PSD: Janis Young, Libby Dechman, and Gerry Ostrove. She is responsible for music, while Mses. Young and

Dechman are responsible for all other subjects. As of June 20, 2011, subject and classification lists are now issued monthly and are intended to increase efficiency in the division. The turnaround on proposals, from submission to authority record is now 8-12 weeks. The same average number of proposals will be approved each month. Because of these changes, those subscribed to the MARC Distribution Service may receive files which contain no records.

Subject Proposal System: A new proposal system will be implemented after July 18, 2011 for LCSH, LCGFT, and Children's Subject Headings proposals. The interface will be like that used for LCC proposals. Both LC catalogers and SACO participants will use the same input mechanism and will hopefully allow for increased efficiency at PSD. The new proposal system will automatically check for backdoor headings, that is, broader terms on proposals which have no authority records. The new system will also allow for the elimination of record juggling in Voyager and will reduce the number of manual edits made in producing the Tentative Lists.

Policy Changes: Name and subject headings for Tibet have been revised to conform to AACR2 and LCSH policies at the behest of the scholarly and Tibetan communities. Because the Netherlands Antilles dissolved in October 2010, headings have been made for each of the former constituent parts.

Genre/Form: Over 700 existing authority records have been re-distributed; with the MARC coding changed to indicate the record is a LC Genre/Form Thesaurus (LCGFT) record. Genre/form terms for English language law resources were implemented on June 15, 2011. The joint LC/MLA genre/form music project is developing the syndetic structure for the first 800-1000 terms. LC is continuing to work with the American Theological Library Association (ATLA) on the religion genre/form project. ATLA will present a thesaurus to PSD later this year. The literature project will be beginning soon.

Questions: When will general genre/form terms be included in LCGFT? General genre/form terms will be added at some time in the future, but no date has been set. Is LC working with Ex Libris to display GF terms better in their OPAC? Yes, LC is working with Ex Libris on this problem.

2. Authorities as Things – by Karen Coyle, librarian and freelance consultant

The term authority control is bandied about, but what exactly does it control? According to Lois Mai Chan, authority control is the mechanism by which consistency is ensured for headings that stand-in for the same person or subject. What this means to Ms. Coyle is that the focus of authority control is on the display of the name, which is completely different from the Semantic, or Data Web. Authority control emphasizes the correct display of a name, etc.; the data web emphasizes the "thing" behind the name. Statements can be said about the "thing" and other "things" can be said about those things. For example, the "thing" William Shakespeare, has a name, has a place of birth (Stratford-upon-Avon), was married to Anne Hathaway and wrote the play *As You Like It*. Other "things" can be said about those "things": Stratford is a part of Warwickshire, Anne Hathaway was born in 1555, and *As You*

Like It is a comedy. Library authority records are just concerned about the name display. The data entered in an authority record is there to establish the form of name. But this emphasis on the form of name obscures the "thing" behind the name.

To get closer to a data model which describes "things", they first must have identifiers which are consistent, persistent, and unique within some context. Consistence means that the same identifier must always mean the same thing. Persistence means that the identifier will stay the same, forever. Since identifiers must be unique, the concept it represents must also be defined narrowly, which leads to more specificity than is possible when using natural language terms. Identifiers also allow for internationalization, in the sense that terms in different languages which represent the same concept can be brought together under one identifier. An example of this is the VIAF project. Another benefit of using identifiers, especially on the Web, is that different identifiers which represent the same thing can be linked together, using tools of the Semantic Web.

The FRBR model is very much in the spirit of the Semantic Web. FRBR has entities, or "things", and these "things" have relationships, or "links", to one another. On the other hand, Coyle views the FRAD/FRSAD model, which describes authority data, as still too heading oriented, since it defines the name of a person, concept, etc. as a "thing" in addition to defining the person, concept, etc. itself as a "thing". It would be better if the name of a "thing" were considered more as a label for a person, concept, etc. instead of a "thing" in its own right.

In conclusion, authority control in the future will control the names of "things" by using unique identifiers to represent the "thing" and by creating labels for those "things".

Questions: One of the things that you stress on your Web site is that libraries and librarians should be interested in creating data sets of things. Is the creation of "things" the direction you think authority control should be going? In a sense, yes, because the creation of labels is not going to be that useful in a machine environment. Labels are for users, but computers need more with which to work.

3. Authority Control, New Library Standards, and the Semantic Web – by Gordon Dunsire, librarian and freelance consultant

RDA is optimized for a scenario 1 implementation, that is, each of the FRBR entities are represented by a record. These records are then linked to each other by the relationships they have with one another. Mr. Dunsire hints that libraries might want to be aiming for another implementation scenario beyond this one, which he names "scenario 0". But before getting to this idea, the current state of authority control in libraries must be examined.

In the current environment, authority control is done by matching the name string on the bibliographic record with the name string in the authority record. This is inefficient in the sense that if the heading string changes on the authority record, all of the strings in the bibliographic file must also be changed. Instead of doing this, an unchanging identifier should be used that would link the records to one another. The natural language form of the

name would still be displayed to users, but the link using the identifier would be in the background, allowing the records to stay connected, even if the display form of the name changes. An identifier can be used for anything which represents a real-world object or concept.

RDA extends this possible use of identifiers to “stand-in” for controlled terms. Identifiers could be used for each term that is used in the Content Type vocabulary. The use of identifiers in library data could also go further by having authorities of authorities. For example, identifiers could be used for a place of birth in an authority record for a person. Identifiers allow for the safe linking of pieces of data and can be continuously linked. But once those links reach a natural-language (or literal) value, the linking must stop. The idea is to use identifiers as much as possible until it is no longer possible to do so.

The concept of linked data is not a new idea and has been around since the start of authority control. The Semantic Web and linked data is simply an extension of the concepts of authority control. The advantages of this model is that something that is identified only needs to be created once, but it can be linked many times over. The display of the data no longer needs to be the thing that most concerns librarians. An authority record need no longer contain just data related to the form of a name of an entity, but may also include linked data to other bits of information about that entity. If this idea of linking atoms of data to one another is taken to its logical conclusion, then this will lead to the disappearance of bibliographic and authority records as they are known today. Instead bibliographic control systems will contain simple, interrelated statements which can be processed and manipulated much more easily by machines.

To get to this scenario of easily processed machine data, it must be modeled using the Resource Description Framework, or RDF. It is designed for the machine processing of metadata at a global scale on the Semantic Web. But because machines are simple and dumb, everything must be disambiguated and done in a very simple manner. Everything must be reduced to a simple statement in 3 parts, called a triple. The first part of the triple is called the subject, i.e., what the triple is about. The 2nd part is called the predicate and indicates the nature of the statement. The 3rd part is the value of the statement and is called the object. For example, the natural language statement, “The title of this book is ‘Cataloging is fun!’” can be re-stated in a triple as Book [subject] has title [predicate] “Cataloging is fun!” [object].

The trick to making a triple machine actionable is that there must be an unambiguous way of identifying each part of the triple. Human language labels are insufficient for this task, because of the slippery nature of language. One way to overcome this on the Web is to use a Uniform Resource Identifier, or URI. At its most useful, it is in the form of a URL, something which machines process on a regular basis. A URI has no intrinsic meaning, it is just an identifier. While a URI may look like a URL, the URI does not necessarily need to lead to a Web page. URIs which do lead to human-readable Web pages are just providing value-added information. URIs that return human readable information are called dereferenceable URIs. RDF requires that the subject and predicate portions of a triple to be URIs. The object can be a URI or may be a literal string. In RDF, predicates are called

properties, and can be thought of as the verb of a triple statement. Properties link specific instances of two things. Properties are the links in linked data and are the pathways through the Semantic Web that lead to human-readable metadata.

Catalog headings can be managed in the same way as other controlled vocabularies. They can serve as natural language labels for RDF entities. By using global identifiers, such as URIs, to identify specific objects and RDF to relate objects to one another, it allows for distributed authority control. Because authority control would be on the Web, there would be no need to copy and maintain authority records in local systems. If a heading is needed on a bibliographic record, all that is really needed is a link out to the identifier for that heading. Because the linking is done on the identifier level, different labels, i.e. headings, for the same thing can be linked to that identifier.

Questions:

Q: Since the mechanisms exist already for linking data, what is stopping libraries from using the architecture of the Semantic web to describe the objects we own right now? **A:** The thing that is stopping us are the lack of properties available on the Semantic Web. Until more properties exist with URI identifiers, things cannot move forward. This is a difficult process, but things are reaching a critical mass.

Q: Are ILS vendors talking to the Semantic Web people? **A:** It doesn't appear that there is much talking going on, but in order to take full advantage of the Semantic Web, a complete re-thinking and conceptualization will need to take place on the part of libraries and vendors.

Q: What can libraries be doing right now to help this linked data revolution along? **A:** Start talking to colleagues and managers about the concepts of the Semantic Web to make it understandable to them, to show them why it is a good idea.

Q: Could you explain the difference between a URI and URL again? **A:** The URI may look like a URL, but it does not need to take one to a normal Web page. URIs are simply identifiers for objects or concepts. URIs which do display Web pages with human-readable data are doing so because the owner of the URI has set up a URI de-referencable service which re-directs the URI to something which humans can consume. This is essentially a value-added feature for the URI and is not essential.

Q: There seems to be a hall of mirrors aspect to all of this. How can you be sure that the machine is processing "correctly" and the data is correct? **A:** Libraries will need to increasingly rely on provenance of information when evaluating information resources.

4. The Getty Vocabularies: Issues in Authority Control for Art and Architecture – by Robin Johnson, senior standards editor in the Getty Vocabulary Program, Getty Research Institute.

The Getty Vocabularies describe art and architecture objects using "non-authoritarian" authorities. Persons who use the vocabularies can make use of the richness of the

vocabularies without having to be tied to a preferred term. The vocabularies are freely available to anyone to lookup on the Web. The vocabularies are also available for download in XML. Those who wish to make extensive use of the vocabularies should seek a license from the Getty Research Institute.

No terms are preferred in the vocabularies. There are terms marked “preferred” in the vocabularies, based on literary warrant, but users of the vocabularies may choose from any of the terms listed within a vocabulary record. The Getty vocabularies are thesauri. There are three main types of relationships: equivalence (same concept), genus-species (broader, narrower, and related terms), and associative (e.g., material, medium, etc.).

There are 4 Getty vocabularies. The Union List of Artists Names or ULAN, is somewhat equivalent to a name authority record file. For the artists in the database there are dates of birth and death, field of activity, roles, nationalities, forms of name, and associated people and corporate bodies. The scope of ULAN is from the antiquity to the present and includes artists, galleries, sitters, etc. In the future, publishers may also be added to ULAN. An emphasis is being made to add non-Western artists.

The Getty Thesaurus of Geographic Names, or TGN, is a structured vocabulary for places, structured on whole/part relationships. The coverage is from pre-history to the present and includes extraterrestrial places, such as planets, moons, etc., geographic areas and formations, and political divisions.

The AAT, or Art and Architecture Thesaurus, is a structured, hierarchical, vocabulary of art and architecture concepts. The scope is from pre-history to the present with no geographical limitations. Currently more qualifiers are being added to terms because of the addition of multi-lingual terms. Legacy data is also being updated or removed. Conservation terms are also being added.

The newest vocabulary is the Cultural Objects Name Authority (CONA). The focus of the vocabulary is the work of art itself. CONA is the first Getty thesaurus that will have links to the terms in the other Getty vocabularies and that can link out to other vocabularies. CONA is currently in the beta phase and will be released in either late 2011 or 2012.

One of the most exciting additions to the vocabularies in the past year has been multilingual terms. A single concept, object, etc. can be searched using various language forms. While the vocabularies had included some translated terms, this is the first time a systematic effort has been made to add these types of terms. The hope is to have the thesauri available for use in multiple languages. The Getty is also participating in the VIAF project and is working on making the vocabularies available in the SKOS format.

Business Meeting

Caroline Miller, University of California Los Angeles will be the new Vice Chair/Chair Elect. Robert Rohrbacher, Stanford University, Bob Wolverton, Mississippi State University, Karen Miller, Northwestern University, and Elizabeth Bodian, Chicago Public Library, are the new

Members-at-Large. As all were running unopposed, they were elected by acclamation. There were two candidates for Secretary: Nate Cothran, Backstage Library Works and Mark Scharff, Washington University. A vote was held and Scharff was re-elected to the position of Secretary. Lynnette Fields, Southern Illinois University will become the past chair, with Melanie McGurr, Ohio State University as the new chair.

Gordon Dunsire's and Karen Coyle's presentations will be available soon on the ACIG portion of the ALA Connect Web site at <http://connect.ala.org/node/65335>. Robin Johnson's presentation may not be available due to copyright restrictions. She must first talk to the Getty Research Institute to obtain permission for posting on the Web site.

The rest of the discussion centered on the evaluation of the program and ideas for the 2012 Midwinter and Annual meetings. These included the future role of authority librarians, function vs. form, and preparing vocabularies for the Semantic Web.

Online Audiovisual Catalogers (OLAC) Cataloging Policy Committee (CAPC) Meeting Friday, June 24, 7:00-9:00PM

Announcements

- Heidi Frank and Susan Wynne are rotating off of CAPC. Anchalee (Joy) Panigabutra-Roberts, Diane Robson, and Stacie Traill will be rotating onto the committee. Dawn Loomis and Katia Strieck are the new interns.

Reports and Discussions

- MARBI Report - given by Cate Gerhart, MARBI Liaison, University of Washington. The following proposals for the MARC formats generated discussion:
 - *Proposal No. 2011-02*: This proposal would provide more granularity in the MARC formats to code publication, distribution, and manufacturer information. The 1st option would be to define a new 264 field and to use indicators for the function (production, publication, distribution, or manufacture). The 2nd option would be to define separate MARC fields for each function. After a straw-poll of CAPC members, CAPC preferred the first option.
 - *Proposal No. 2011-03*: This proposal would allow for the specific coding of a copyright date. Option 1 would propose a yet-to-be-determined new field for the copyright, using \$a for © dates and \$b for (p) dates. Option 2, instead of using subfield codes, would use indicator codes for the two types of dates. Option 3 would make no distinction between © and (p) dates. Discussion immediately dismissed using option 3. After some further discussion, CAPC preferred option 2.
 - *Proposal No. 2011-04*: This proposal would extend the use of the 377 field to include language of the expression and would add this field to the bibliographic format. It is currently defined as the associated language of a person or corporate body. Concerns were raised about why a new field was not being proposed for this attribute. It was explained that the properties of the field would be deduced from the entity described in the record.
 - *Proposal No. 2011-08*: This proposal would allow for the coding of additional carrier characteristics. Option 1 would expand the existing 340 field to allow for additional coding of some carrier characteristics. This option would also define a field 344 for sound characteristics, field 345 for projected media characteristics, field 346 for video characteristics, and field 347 for digital file characteristics. Option 2 would replace fields 344-347 with a single field that would include individual subfields for the carrier characteristics. In both of these options, carrier characteristics would also be able to be in a repeatable \$b in the 300 field and could also be coded in a 500 field, using \$i to denote the type of characteristic. Discussion was more in favor of using option 2, since the display to users would be more useful and it was felt that the data would be more machine-actionable. Some dubiousness was expressed with the proposal to use \$i in the 500 field and the repeatable \$b in the 300 field, especially as concerns granularity and any mappings to future metadata schemas.

- For further details of other MARC proposals and discussion papers, please see the MARBI liaison report from Bruce Evans, MARC Subcommittee Chair, BCC.
- CC:DA Report – given by Kelley McGrath, CC:DA liaison, University of Oregon. For details of the CC:DA meeting at ALA, please see the report from Mark Scharff, Descriptive Cataloging Subcommittee Chair, BCC.
- Video Language Coding Best Practices Task Force – given by Kelley McGrath. At ALA Midwinter in San Diego, the task force proposed in MARC Proposal No. 2011-01 to redefine \$h of the 041 to represent the original language of the work, even if the resource is not in translation, and to define a new \$k for intermediate translations and a \$m for the original language of accompanying material. This proposal, with amendments, was accepted by MARBI. This means the recommendations of the Task Force have now all been approved, however some of the changes have not yet been implemented in the MARC formats. Once this happens, the draft of the best practices document will be updated and sent to CPAC for approval.
- LC Genre/Form Headings for Moving Images Best Practices Task Force – given by Susan Wynne, University of Wyoming for Scott Dutkiewicz, Clemson University. Examples in the document have been updated to reflect new MARC coding guidelines from the Policy and Standards Division, Library of Congress. Since the draft has already received a public airing and the changes resulting from it were minimal, the document will be submitted to CAPC for approval.
- Audiovisual Materials Glossary Update Task Force – given by Heidi Frank, New York University. A public interface for the glossary has been developed and is available on the OLAC Web site at <http://www.olacinc.org/AV-Glossary/>. The interface is still under development. The majority of the terms from the original A-V Glossary publication by Nancy Olson have been entered in the database. The task force is compiling lists of new terms to add to the glossary. Once the list is complete, it will be evaluated against the glossary to avoid duplication, overlap, or synonymous terms. A new introduction is also being drafted. The task force feels that someone or something will need to be responsible for updating and revising the Web site over time.
- RDA Testing – given by Kelley McGrath. The work of the OLAC-MLA funnel is now officially finished. Ms. McGrath thanked all of the participants for their hard work. The committee expressed its appreciation for all of the effort she put into making the funnel work.
- Library of Congress Report – given by Janis Young, Library of Congress.
 - On behalf of Andrea Leigh of the Moving Image Section of the Library of Congress, Ms. Young presented the committee with guidelines the section will be following in copy-cataloging DVD and Blu-Ray discs. All copy will be changed in the following ways: coding in 008/15-17 will be based on the 257 field instead of the 260 \$a and 008/33 will change the type from "v" to "m". The GMD in 245h will be removed. A 257 field will be added to the record and will contain the country of production. An interim decision has also been made to code the main work contained in a DVD or Blu-Ray as an added entry instead of as main entry. Because these changes are being made to accommodate AMIM2 practice, there was some concern whether LC's copy-cataloging would overlay AACR2 records in OCLC. No answer was able to be provided.

- For genre/form details, please see the report from Hermine Vermeij, Subject Access Subcommittee Chair, BCC.

Old Business

- BIBCO Standard Record for Moving Image Materials, Graphic Materials, and Electronic Resources – given by Walt Walker, Loyola Marymount University. Recommendations of the task force to PCC, based on their original charge, are complete. The task force was then asked by PCC to develop provider-neutral guidelines for online moving images and graphic materials. Those guidelines have also been completed. The task force expects to be contacted by PCC to develop guidelines using RDA.
- Moving Image Work Grant – given by Kelley McGrath. Ms. McGrath has applied for an internal grant at the University of Oregon to extract work record data from MARC records using the XC Metadata Toolkit. It is unknown at this time whether the grant has been accepted. There will be a need for volunteers to help with the project.

New Business

- After the testing of RDA, it was initially thought that a RDA Examples Task Force should be formed to add RDA examples to current OLAC documentation. After some e-mail discussion prior to this CAPC meeting, it was decided that it might be better to do wholesale revisions of certain important documents. Documents will be identified and volunteers from the OLAC membership as a whole will be solicited to work on each document after ALA Annual.