

**BCC2010/MD/1****Report from the ALCTS Metadata Interest Group  
and other metadata-related meetings  
Jenn Riley, Chair, Metadata Subcommittee, MLA BCC**

**ALCTS Metadata Interest Group** (Formerly Networked Resources and Metadata Interest Group) Annual Meeting,  
January 17, 2010, 8:00-10:00 AM

The ALCTS Networked Resources and Metadata Interest Group has become the Metadata Interest Group (MIG), effective at ALA Annual 2009. The current Chair is well-known to MLA members: Brad Eden, Associate University Librarian for Technical Services and Scholarly Communication, University of California, Santa Barbara. The MIG meeting at ALA Midwinter 2010 featured two presentations from leaders in the metadata community, followed by a brief business meeting.

In the first presentation, Jennifer Bowen of the University of Rochester eXtensible Catalog (XC) project reported on the metadata management component of this project. One software package developed by the project is known as the Metadata Services Toolkit. This component allows project staff to set up periodic harvests of metadata from various sources (including library catalogs) via OAI-PMH, provides tools for cleaning up and normalizing this metadata, then re-exposes the enhanced metadata for use in the XC discovery system or any other discovery layer desired. The Metadata Services Toolkit is made up of several components: a normalization service, transformation service, aggregation service, and (planned) authority control service. This software has been released under an open source license and is available for others to implement at their own institutions. Slides from Bowen's presentation to the MIG may be found at [http://www.presentations.ala.org/index.php?title=Image: Bowen\\_ALCTS\\_Metadata\\_Interest\\_Group\\_Midwinter\\_2010.pdf](http://www.presentations.ala.org/index.php?title=Image: Bowen_ALCTS_Metadata_Interest_Group_Midwinter_2010.pdf). Bowen also gave presentations on other aspects of the XC project at other ALA Midwinter 2010 events, including to the FRBR Interest Group and the Authority Control Interest Group (ACIG). The eXtensible Catalog project home page, which includes links to download the open source XC toolkit, is <http://www.extensiblecatalog.org>.

In the second presentation, Roy Tennant of OCLC Research reported on informal work he has performed towards a "Massive Metadata Mashup." OCLC's Cloud Library project is investigating the intersection of materials in offsite storage and digitized content to see if there are operational efficiencies to be found for libraries. Tennant's work involved downloading the freely-available Hathi Trust metadata records, using the OCLC XID service to enhance these records with OCLC numbers, then merge these enhanced records with WorldCat records and data on holdings in storage facilities to gain an overall picture of the overlap. While the project lessons overall are interesting, to the MIG audience, the mechanisms to perform this work are perhaps more important. Tennant presented three lessons learned related to metadata issues: identifiers are essential when dealing with large metadata sets, sometimes you need to go outside of standards to move forward with a project, and when dealing with large data processing activities, small incorrect assumptions or mistakes can have large consequences. Slides from Tennant's presentation to the MIG may be found at [http://presentations.ala.org/index.php?title=Image: RoyTennant\\_metadata\\_compressed.pdf](http://presentations.ala.org/index.php?title=Image: RoyTennant_metadata_compressed.pdf).

The Metadata Interest Group business meeting was brief, and primarily focused on reports from liaisons. Jenn Riley, MLA liaison to MIG, reported on MLA activities in the business meeting, including the work of the MLA BCC Metadata Subcommittee in partnership with the ALA Association for Library Collections and Technical Services (ALCTS) Preservation and Reformatting Section (PARS) Audio Metadata Task Force to create an online resource explaining the use of a variety of metadata standards related to digital audio, and metadata-related sessions to be presented at the upcoming MLA 2010 Annual Meeting in San Diego. Among the business discussed of interest to the MLA community was an ALCTS preconference to ALA Annual 2010 on June 25, 2010, covering the XSLT technology: <http://www.ala.org/ala/mgrps/divs/alcts/confevents/upcoming/ala/ac10/sxlt.cfm>, and a session the MIG will host at ALA Annual 2010 on projects spanning the divisions between libraries, archives, and museums.

The Metadata Interest Group maintains a blog at which summaries and links to presentation slides from other ALA Midwinter 2010 meetings and presentations relevant to metadata issues are listed. This blog may be found at <http://blogs.ala.org/nrmig.php>.

## **ALCTS Preservation and Reformatting Section (PARS), Audio Metadata Task Force**

The BCC Metadata Subcommittee is partnering with the ALCTS PARS Audio Metadata Task Force to create an online resource helping implementers to select appropriate metadata standards for digital audio projects. The Task Force did not have a formal meeting at Midwinter, but several representatives did meet informally to plan a strategy for completing the project by February 2010. This informal meeting was also attended by Charles Kolb, a program officer at the National Endowment for the Humanities (NEH), who expressed interest in endorsing the resource developed for use by its grantees.

## **Other ALA Midwinter 2010 Sessions of Interest to MLA Members in the Area of Metadata**

The FRBR Interest Group met Friday, January 15, from 10:30 AM – Noon. Discussions from this session of interest to the MLA metadata community include:

The eXtensible Catalog (XC) project's early decisions to display search results always as Manifestations, with data elements related to search terms from other entities, rather than displaying the entity with the term that matched the user's query; and their planning for "hourglass" displays (grouped around a Manifestation) and "pyramid" displays (grouped around a Work).

IFLA plans to embark on work to harmonize the FRBR, FRAD, and FRSAD models once the final version of FRSAD is published.

The European Library (TEL) project has posted online a prototype FRBRized system of literature written by Nobel Prize winners. The system may be found at <http://digmap2.ist.utl.pt:8080/lgteFrbr/indexFrbrClustering.jsp>.

At the Cataloging and Classification Section (CCS) Forum, two presentations are potentially of interest to the MLA metadata community. First is the presentation "Could we Create a Semantic Web Data Model for Subject Cataloging?" by Martha Yee, Cataloging Supervisor at the UCLA Film & Television Archive. Yee's research is currently focusing on how library metadata might be represented in the context of the Semantic Web, demonstrating both what she has been able to model in this way, and potential barriers as well. Powerpoint slides for this presentation may be found at <http://myee.bol.ucla.edu/Boston%20powerpoint.ppt>. The second relevant presentation was by Tina Gross, Daniel N. Joudrey, and Arlene G. Taylor, on "The Effect of Controlled Vocabulary on Keyword Searching." This project is a follow-up to a 2005 study published in *College & Research Libraries* assessing what proportion of searches performed would fail to match if controlled subject headings were not applied to records. The new study was performed on a catalog that had been enhanced with table of contents data since the first study was performed, which is commonly cited as one method to improve search results. The new study has resulted in a preliminary conclusion that for English-language materials nearly 25% of searches match only data in controlled subject headings, down from just over 35% in the first study.

One additional relevant presentation at ALA Midwinter was hosted by the Preservation and Reformatting Section (PARS) Intellectual Access to Preservation Metadata Interest Group. Lisa Schmidt, Electronic Records Archivist at Michigan State University, discussed digital curation planning at MSU. Work done so far includes a survey of units on campus that manage digital content, to understand their current metadata and data curation practices, which will push the development of specifications for campus-wide curation services. Presentation slides may be found at <http://msudcp.archives.msu.edu/wp-content/uploads/2010/01/ALA-Midwinter-2010.ppt3.pdf>.

*Submitted by Jenn Riley*

*Last updated March 11, 2010*

