

Report on the session, “Workflow Design for Metadata Creation,” sponsored by the Bibliographic Control Committee, Metadata Subcommittee

The **Metadata Subcommittee** is new, having been formed at the 2009 annual meeting in Chicago with Jenn Riley (Indiana University) as chair and nine other members. On Wednesday, March 24, 2010, the subcommittee presented a program on “**Workflow Design and Metadata Creation**” at the Music Library Association (MLA) annual meeting in San Diego, California. In this program, the three presenters, Caitlin Hunter (Library of Congress), Amanda Harlan (Baylor University), and Renée McBride (University of North Carolina-Chapel Hill), described very different metadata projects at their respective institutions and the problems and successes they encountered.

After a brief, informal announcement of a vacancy on the subcommittee, subcommittee chair and session moderator **Jenn Riley** opened the session with some reflections on the many basic decisions that need to be made when creating any kind of metadata for a project. With these questions as a backdrop to the session as a whole, Riley introduced each presenter in turn.

Caitlin Hunter began her presentation, “**MAVIS Audio Collection Management Software and the National Jukebox Project,**” with a description of the organization and staffing at the Recorded Sound Section of the Motion Picture, Broadcasting, and Recorded Sound Division at the Library of Congress, where she is Processing Unit Head. She then described how the Division uses MAVIS (Merged Audio Visual Information System), a collection management software program developed by ScreenSound in Australia, as an inventory control tool in the preservation and archiving of sound and moving image material. At LC, a technician creates metadata records in MAVIS containing the title, component (copy) and carrier information, which the cataloger then reviews. The system generates pull lists; recording engineers can also add metadata, since they must listen to the recordings in the process of digitizing them. A workflow software that can “sit” on top of MAVIS is still in development.

Ms. Hunter then gave a report on the **National Jukebox Project**, in which the Library of Congress is partnering with the University of California-Santa Barbara and Sony Corporation to make thousands of sound recordings from the 78-rpm era available for listening online, particularly those on the Victor and Columbia labels. This project thus requires coordination of participants on opposite coasts. On the LC side, a technician prepares a spreadsheet with representative samples. Pull lists are created before the recordings are digitized; technicians compare copies of discs and select the one in the best condition for digitization. A similar workflow is followed at UC-Santa Barbara, which is contributing much discographical information. The digitization on the West Coast is being done by a Seattle-based company, Reclaim Media.

Amanda Harlan, Metadata & Catalog Librarian at Baylor University, gave an overview of the **Frances G. Spencer Collection of Popular Sheet Music and the Black Gospel Restoration Project** and outlined the procedures used to create metadata for them. The sheet music collection contains over 30,000 titles from the late 18th to early 20th centuries; digitization began in 2000. Shelf cards and cover images are outsourced to the Flourish music cataloging service. The gospel music collection, on the other hand, consists of sound recordings in various formats from the 1940’s through the 1970’s; a professor at Baylor collaborated with a wealthy businessman who wanted to help preserve this music. So far, 2,786 pieces of sheet music and 1190 albums of black gospel music have been cataloged.

In creating metadata for these collections, Ms. Harlan recommends going from standard having richer content (such as MARC or MODS) to simpler one (such as Dublin Core), rather than the reverse. She begins with the MARC record and uses MARC Edit to extract Dublin Core (for CONTENTdm) and MODS records. MODS is used for preservation metadata; Ms. Harlan mentioned that it is also used in other venues for “dark archives.” For technical metadata, she uses JHOVE (which is compliant with the NISO Z39.87 standard) for images and AES-X098B for audio. AES-0X098C is used for digital provenance and process history (administrative metadata), and METS for structural metadata. Ms. Harlan noted that METS is more flexible than PREMIS. Rights metadata schema for the Black Gospel Music Project is taken from several sources and combined into a single schema called the Baylor Copyright Metadata Schema (BCMS). This is not needed for the sheet music, as most of that collection is in the public domain.

These collections can be accessed at this link: <http://contentdm.baylor.edu/>

Renée McBride (Head, Special Formats & Metadata Section, Resource Description & Management Dept., Davis Library, University of North Carolina-Chapel Hill) addressed the all too common scenario of projects begun and left unfinished in “**Look What We Got! How Inherited Data Drives Decision-Making.**” She began with a historical synopsis of the fortunes of a collection 129 bound volumes of 19th century sheet music at her library, consisting of nearly 3,500 pieces. Over the years several attempts were made to catalog this collection and, in the mid-1990’s and again in 2004, even to digitize it and put it online, but each time the project was interrupted for one reason or other, usually lack of money or staffing. In 2008, Library staff and students enrolled in the School of Information & Library Science began transferring the collection from MySQL, an open source database software, to CONTENTdm. As a result of this on-and-off approach, there has been great inconsistency in the “legacy” metadata. Ms. McBride detailed how she reduced the number of fields in the Dublin Core records, and created a “data dictionary” of controlled vocabulary and used CONTENTdm’s “find and replace” function to clean up the database. Internal consistency is the immediate goal; conformity with established Library of Congress authorities may be pursued later.

The 19th Century American Sheet Music project can be found at:
<http://www.lib.unc.edu/dc/sheetmusic/>

After the presentations there was a brief question-and-answer period. One discussion concerned collaboration with public services and systems librarians in metadata creation. Ms. McBride commented that she kept end users in mind when cleaning up her database, and that CONTENTdm requires you to choose which fields are searchable and which ones are visible to the public. Responding to a question about the Black Gospel Music project, Ms. Harlan said that Baylor does not physically own all of the recordings, and that some of them are reproductions.

Submitted by Felicia Piscitelli, Texas A&M University