

Catalogers Group
March 9, 2006 Minutes

Present: Melissa Beck, Sharon Benamou, Valerie Bross, Claudia Horning, Chamya Kincy, Sara Layne (presenter/moderator), Caroline Miller, Nancy Norris, Hao Phan, Laura Horwitz, Louise Ratliff, John Riemer, Gia Aivazian (Recorder).

Topic: Discussion of issues that arose at two conferences last month:

A. Music Library Association Annual Conference, Memphis

<http://exlibris.memphis.edu/music/mla/>

Sharon Benamou reported on the discussion on Genre/Form headings at a program sponsored by the Technical Services Roundtable. She then concentrated on three of the four speakers. Geraldine Ostrove of LC who had sent out a questionnaire on MLA-L and possibly listservs to solicit feedback about the different ways institutions use genre and form headings, said that feedback had been low. There was no consistent use of genre headings among the different types of institutions be they public libraries, large or small academic libraries. Subject headings for music are primarily forms or forms plus medium of performance and are currently coded as 650s. Since LC is planning implementation of 655 and the creation of 155 authorized headings, Ostrove made clear that MLA needs to be involved in the implementation. She discussed the methods LC was using to determine which headings needed to be flipped, including left-anchored searches on names of instruments followed by “and”, “choir”, etc.

More important was the need to address conversion of complex or ambiguous subject strings. Some terms express both “of”ness and “about”ness but do we need duplicate records? How do we create syndetic structure for terms used as both form and topical subject? Some strings only show themselves to be form headings in a subdivision. What happens to the tagging for these? Examples: --Hymns, --Musical settings, --Songs and music. Some examples were discussed.

Rules will need to be created for assigning different or same topical and form headings covering the same general concept to the same manifestation.

Do we need to break apart strings under certain circumstances?

Robert Maxwell spoke of the need for genre and form headings. BYU creates authority records for genre terms and has in the OPAC a separate search button for the genre and form headings for literature, but not music. BYU is a SIRSI library and one may view its catalog at: <http://www.lib.byu.edu/>. He mentioned that in the Subject Subcommittee of the Bibliographic Control Committee a

discussion took place about creating authority records for all form and medium of performance headings. Currently, although Joan LoPear told Sharon that this was not true in the past, medium of performance is assigned according to rules in SCM H 1917.5 but they are usually not established and there is no requirement for establishment. One example Sharon had from this week was Concertos (Marimba). Rules for order of instrument groups, order of instruments within families, use of “and” and “with.”

The third speaker was OCLC-based Becky Dean who talked about FAST (Faceted Access to Subject Terminology)

<http://www.oclc.org/research/projects/fast/> and the OCLC Terminologies Pilot

<http://0->

www.oclc.org.library.lanec.edu/productworks/terminologiespilot/browser.htm .

She explained that music subjects are problematic when attempt is made to slot portions of the subject strings into the designated facets of topical, geographical, etc. The basic problem is the use of forms, such as “operas” and “sonatas” that are tagged as topical headings. Many of these have medium of performance added on but these would be left out of FAST. She felt that LCSH rules were antiquated and too complex, but so far there is no better solution.

B. Webwise Conference, Los Angeles

<http://www.getty.edu/webwise2006/conference> .

Sara then invited Luiz to share with the group one of the topics covered at this conference, named Automated Metadata. Luiz informed the group that this conference had lasted two-and-a-half days, including one pre-conference day, and that four or five from UCLA had attended. The conference essentially was about metadata creation and use. The segment Luiz spoke about was on Automated Metadata Generation. The questions posed were: Is the right question “To automate or not to automate metadata,” or “How to automate metadata?” Should it be fully-automated or semi-automated? Two presentations related to this topic were:

- Steve Mitchell (UCR) *Automated Metadata Generation and New Resource Discovery Software and Services*. (Available at <http://datafountains.ucr.edu/IMLS2006presentation.ppt>)

- Victoria McCargar. *Newspapers Online*.

Mitchell presented new software developed for both generation and extraction of Metadata: iVia and Data Fountains. The goal here is to create tools to assist experts reduce routine work. As automated tools, iVia and Data Fountains generate metadata, including assignment of LCC (classification) and LCSH based on specific algorithms developed.

iVia
Data Fountains

<http://ivia.ucr.edu/>
<http://datafountains.ucr.edu/>

McCargar's presentation demonstrated the more radical approach of the newspaper industry which is eliminating human indexing in favor of a complete automated approach to indexing newspapers online. Here, there seems to be a problem of which way to go, costs being part of the problem. John asked about the use of controlled vocabulary. Sara offered that a machine could generate a controlled vocabulary assignments if fed 200 examples correctly cataloged.

The next presentation was by Chamyra on a topic within the larger topic of issues and challenges concerning digital projects. The specific subtopic within this panel that attracted Chamyra's attention was one that described a project currently underway at the Texas Center for Digital Knowledge, located at the University of North Texas. This is the **MARC Content Designation Utilization Project** (or MCDU) <http://www.mcd�.unt.edu/?p=30>.

The panelist representing the project was one William Moen. Even though the main panel focused on digital projects in general, the speaker related his particular project of MARC utilization *within* that broader context by ascribing the concept of **artifact** to the MARC record. He said that MARC records are artifacts that can be used for inferring the decisions and policies of a "cataloging enterprise." And, that analyzing these "artifacts" can help measure the utilization of metadata within an institution which can assist aggregators working with digital repositories.

Moen then went on to describe the actual project where about 7.6 **million** LC-created records were examined. Now, apparently there are almost **2,000** fields and subfields in the entire MARC format. But, according to the findings of the MCDU project, only **14** of these 2,000 fields occurred in **90%** of the records. It was also found that only **21** of the fields/subfields occurred in **80%** of the records. Also, another thing that they found was that as many as **110** fields/subfields occurred in less than **1%** of all records [possibly implying that the format is too complex and there is a surplus of unnecessary elements???

In the future, the MCDU project is proposing further initiatives that build upon this initial analysis. Some of these initiatives include the development of a database system that could analyze trends and patterns of MARC field utilization. They also want to be able to develop methods that will help them systematically **identify "core" elements** in bibliographic records based on **occurrence**. And they also want to examine standards for core elements in existing schemes [meaning DLC's Core level 4?]. In the long run, they hope to develop a **methodology** that will provide an **empirical basis** for understanding the decisions about core elements in metadata schemes. The hope is that library/information organizations will be able to use these methodologies and tools to analyze **local**

utilization trends. Some of the implications of this effort could possibly be contributions to changes in cataloging rules, practices, policies, and standards.

Sara how we thought this type of research could be put to use. Seldom-used elements might be candidates for elimination from the MARC format. Heavily used elements, e.g. the Literary Form byte in the fixed field (008/33; 006/16) might warrant adoption of finer breakdowns.

Louise, then, was invited to speak of the **Electronic Cultural Atlas Initiative** <http://ecai.org>, which is about how to link and relate various pieces of metadata and combine them in a new way. Louise said the presenter was Michael Buckland, Co-Director of ECAI at UCB. This initiative, Buckland informed the attendees, had its beginnings in 1997 in Berkeley, where an informal meeting of scholars from several countries established the ECAI to improve understanding of history and cultures through increased attention to both time and space... Time and location are universal attributes for sharing cultural knowledge across different languages, disciplines and points of view, he said. Time and location can be used to integrate data from museums, libraries, heritage sites, and research projects around the world. Thus, the mission of ECAI is *to advance scholarship through increased attention to place and time*. Louise said there were a dozen projects that looked interesting. To make a small demonstration on screen, she picked the Silk Road atlas to test the use of time and place (the what, where – both place and space, and when through the use of the ECAI Metadata Infrastructure, i.e. the various resources) with maps, including animated maps. She thought the project was a powerful argument for the use of LCSH.

There followed some discussion on the merits of this very attractive project but also doubts were expressed that it could possibly succeed in being all-encompassing in terms of all languages, all events at all points in time and space.

Gia Aivazian