MOUG's input is valued at OCLC, that the most recent
report on WorldCat Local was noticed at OCLC, and that he will be coordinating a response.

There was a lengthy discussion on the issue of external or third-party content in WorldCat Local, sparked by the "double-edged sword" of AllMusic metadata attached to bibliographic records for sound recordings (in some cases, AllMusic information is richer than the contents notes it replaces, while in other cases the data is missing or meaningless). Third-party or enhanced content is an OCLC priority. OCLC has identified the Internet Movie Database (IMDb) as an additional potential source of enhanced content. Vince asked the audience if IMDb would be a valuable additional potential source of enhanced content. OCLC has identified the Internet Movie Database (IMDb) as an additional potential source of enhanced content. Vince asked the audience if IMDb would be a valuable additional potential source of enhanced content.

MOUG Hot Topics
Jay Weitz, Senior Consulting Database Specialist (OCLC)
Vince Wortman, Product Support Specialist I (OCLC)

Questions and comments at this year's MOUG Hot Topics session were again graciously fielded by OCLC's Jay Weitz, Senior Consulting Database Specialist, and Vince Wortman, Product Support Specialist. While questions focused on the WorldCat Local/WorldCat.org and Connexion products, attendees raised a variety of topics including quality control, duplicate records, input and editing practices, and public record display.

Vince began the session with some good news about WorldCat Local: performer and production notes (fields 508, 511, and 518) will be displayed in the next install. Regarding the status of other changes affecting music, given that music was identified as a specialized area for development at last year's meeting, Vince assured attendees that MOUG's input is valued at OCLC, that the most recent
that it would depend on what exactly the data is. If outside content is rich, yes; if it isn't, no. Jay added that the process of including enhanced data is in its infancy, and as we get to know the data better and work with data providers, we might get a better idea of how to use the data and determine what is manipulable or usable.

A few attendees raised questions on specific searches in WorldCat.org and WorldCat Local. One attendee asked whether a search for a specific edition that returned multiple manifestations in WorldCat.org, but not WorldCat Local, was an intended result of design differences. A second attendee's question addressed a search where recent editions appeared only after "view all editions and formats" was applied but not in a keyword search. Vince answered these questions, noting that the algorithm is in testing and development, and differing effects on the two products is unknown. Regarding the second search, he will pass on examples sent to him.

Again, many questions, concerns, and frustrations about duplicate records surfaced. Citing an astonishing number of duplicate records, historically entered in small numbers by those "damn catalogers" but now in large numbers by OCLC, one attendee asked about what OCLC will be merging to alleviate the problem. Jay then made the shocking confession that in fact he is responsible for the duplicates, declaring, "I sit there every single day, putting those duplicates in, and this is the thanks I get!" On a serious note, he gave a refresher on OCLC's duplicate detection and resolution (DDR) program. In 2004, DDR stopped, and a project began to redo DDR on a new platform. From 2005 through early 2010, it was completely rewritten and expanded to include non-book formats. After testing, the database was processed in order, beginning with record number 1; upon completion in September 2010, over 5 million bibliographic duplicates in all formats were merged. The process also included looking at daily journal files of records that had been added to the database or changed in some significant way. Since early 2010, they have kept pace with DDR. It is run on added records at a one-week lag (not at point of production); the process also runs within foreign-language records. The process is additionally complicated by parallel records appearing as false duplicates.

There was a request for Jay to review the issue of adding a bracketed field 250 to bibliographic records to prevent inappropriate merges. In cases where the only information that differentiates multiple editions appears in quoted notes, the duplicate detection algorithm cannot locate this information. In these cases, catalogers can code this information in accordance with AACR2 1.2B4 that permits a supplied edition statement. A webinar where this is addressed, "Cataloging Defensively: When to Input a New Record in the Age of DDR," is archived and available online.

Another question addressed in the "Cataloging Defensively" webinar is whether a new record correctly input as new could end up deleted or moved to field 019 by DDR. According to Jay, this is true in theory, and is most likely to happen if the Library of Congress enters a duplicate record not caught during the loading process that is then retained over the earlier record. Duplicate detection was designed to codify OCLC's guidelines on when to input a new record and the guidelines in Differences Between, Changes within: Guidelines on When to Create a New Record; it is an automated process but OCLC fine-tunes the algorithm with every reported incorrect merge. Finally, Jay reminded us that the DDR algorithms built into the loading process are only as good as the bibliographic information; even with the best cataloging, bibliographic information is complicated.

Both Vince and Jay shared some of what OCLC is working on related to FRBR, but not yet in any form of WorldCat. OCLC is trying to FRBR-ize the database using the same software developed for DDR to identify records that represent the same thing while getting rid of lesser quality records. Another project, internally called "GLiMiR" (Global Library Manifestation Identifier), is trying to bring together records for a manifestation: instead of multiple records for the same manifestation, users could view one record in their language with all the other records discoverable behind that one. Hopefully this project, still in process, will alleviate many of the problems that have long frustrated music catalogers. Vince added that in WorldCat Local specifically, there work is underway to create "clustering" of results to bring what are actually duplicates together into a representative record. When it comes time to load the algorithm, virtually all records in WorldCat Local will need to be reloaded.

The session closed with a series of questions on cataloging in Connexion. Jay was asked to comment on whether it was possible for OCLC to capture local Connexion edits to incorporate improvements into the master record. Because there is no way to differentiate between good and bad information, there cannot be a way to reliably harvest these changes from daily journals; attendees were advised to edit master records, especially in cases of minimal records. Jay also confirmed that Quality Control welcomes messages requesting fixes in cases where catalogers have discovered
serious problems, such as misidentified works or composers, but do not have the items in hand.

Another question addressed the possibility about expanding capabilities of non-PCC Enhance libraries to edit PCC records. OCLC is talking to the PCC policy committee, with ongoing discussions on the danger of "diluting" the PCC brand if non-PCC libraries can edit a record such that all access points are no longer backed up with authority records. It does look like there will be a change during 2011 to allow edits to PCC records by libraries that do not have BIBCO authorizations but are Enhance and NACO participants. While a conservative addition, this will maintain the PCC brand while expanding editing capabilities.

One Connexion "wish list" item was to limit browse searches to language of cataloging. Jay replied that this is impossible, as that information is not available to the browse index. However, there will be a new version of Connexion Client by early April 2011 that will permit better and new ways to differentiate or qualify by language of cataloging. A cordial disagreement arose among attendees regarding the specific case of cataloging G. Schirmer rental material published with dates in ink on the covers. To some, it seems obvious these dates should be ignored and no new record created. On the other hand, a quoted note could be added to justify the printing date in brackets and the creation of a new record, especially in cases where printing dates are decades apart. The situation is more complicated with printed dates only a single year apart. Jay offered to address this issue if he receives examples. The session's final questions were on RDA. Existing full-level AACR2 records should not be changed to RDA (instances where that may have happened should be reported to OCLC), and on the question of whether a duplicate AACR2 record can be entered if the only existing record is RDA, Jay's answer was short and unambiguous: "No."

Reported by Rebecca Belford, University at Buffalo