Notes from the OCLC Cataloging Community Meeting on 9 June 2023

Session: OCLC Data Quality Update

Presenters

- Cynthia Whitacre, Senior Metadata Operations Manager, Metadata Engagement, OCLC
- Laura Ramsey, Senior Metadata Operations Manager, Data Quality, OCLC

Presentation summary

Laura Ramsey began the presentation with an announcement that merging functionality is coming soon to Record Manager. Announcements and details will appear in the Record Manager and Member Merge Community Centers as soon as the functionality is available. She invited anyone interested in joining the Member Merge Program to contact askqc@oclc.org about joining. She also gave an overview of what the Record Manager merging will look like.

Laura Ramsey went into detail about the DDC enrichment project that is currently in progress. The goal of the project is to add Dewey classification numbers to WorldCat bibliographic records that lack them. These numbers are added programmatically within the same work cluster under strict conditions, and the work was done with Alex Kyrios, Senior Dewey Editor, to ensure correct formatting and conditions. Catalogers can treat these numbers like any other class number in a record, and are free to ignore, edit, or remove if incorrect, as needed. About 33 million records have been processed since February, out of an expected 85 million records.

Another important project in the works is the duplicate detection data modeling. We have collected data on 34,356 pairs of records to train a deduplication model, with Data Quality staff reviewing results and giving feedback throughout the process. This phase is now complete, so the model will be applied to WorldCat in the near future. More detail will be forthcoming, and the April 2023 Virtual AskQC Office Hour has more information about this project overall.

Cynthia Whitacre announced the upcoming schedule of Virtual AskQC Office Hours for the remainder of 2023 and noted that the registration is up for the July session for those that wish to attend. The upcoming sessions are:

- June: Cataloging Rare Materials Defensively, presented by Jay Weitz and Kate James
- July: De-biasing Dewey: Righting the Past by Rewriting the Classification, presented by Alex Kyrios and Kathryn Becker
- August: Rapid Harm Reduction with Locally Preferred Subjects in WorldCat Discovery, presented by Grace McGann
- September: Get Informed About Genre/Form Terms, presented by Kate James
- October: Yet to be determined, so if you have a suggestion, please feel free to submit it to askqc@oclc.org

- November: MARC Fields for Manuscripts and Archival Collections, presented by Kate James

Registrations for these sessions will be posted in the usual places. Also, there will be an OCLC Cataloging Community Meeting in-person at ALA on June 23. It will not repeat this session but cover a few topics and mostly focus on Q&A.

The session concluded with the announcement of Jay Weitz's retirement at the end of June. Cynthia and Nathan shared their experiences working with Jay as well as some pictures from Jay's time at OCLC over the last 41 years. Nathan also shared a few comments from a Kudo board that is now available for everyone to share tributes and memories of working with Jay.

URLs mentioned during the event

- AskQC site: oc.lc/askqc
- Jay's Kudoboard: https://www.kudoboard.com/boards/ZitBYY5P

Audience questions

Q. Does the Dewey project include CONSER serials? Does the deduplication project include CONSER serials?

A. Yes and no. Yes, to the DDC project. We separate those and run very small numbers because of the CONSER distribution impact. We don't want to send 20,000 updates, so we're trickling through those. The data modeling is only dealing with books at this point, so it's not doing serials, it's not doing any other format at this time. We are starting to look at other formats, starting to do some analyzing, but serials will probably be the last—if at all.

Q. For the upcoming project of de-duping MARC records, if a dup record gets deleted, what happens to the library that has that MARC record in their catalog exported from OCLC?

A. The process, when it actually gets to the merging, is not any different than what we do today with DDR. In this process, instead of having DDR go out and find those sets to merge, we're passing those sets to DDR. Everything else happens the same way: we retain the same records we always do, holdings, LHRs, everything gets moved over. The number gets indexed in the 019 field of the retained record; all of that continues to happen.

Also, if the record that you hold gets merged into another record, you can use the WorldCat Updates feature within Collection Manager to get the updated record with the new number.

Q. Is OCLC going to be publishing any of the data modeling real-life findings in a white paper? It might help push public opinion for a particular bibliographic model as being "actually based on data" or as opposed to others.

A. Nathan Putnam will be presenting at IFLA in August in Rotterdam. The presentation will focus on the history and a lot of the information about our machine learning processes, and it will

include this particular project and everything that went into it. We will likely be talking about this for a while.

Q. Are there any plans to remove the 375 (the gender field) from name authority records?

A. There are plans. Over a year ago, the PCC agreed to no longer use this field in authority records for personal names. We have done testing of removal and had offered to LC to remove them. We are just waiting for the Library of Congress Systems Office to give us the go-ahead. And as part of your normal workflow, you can remove them when you come across them. In fact, you are supposed to remove them if you are editing the record for other reasons. And OCLC will notify the community as soon as we are able to start the removal project.

Q. After they are removed, will it be a validation error if people try to add a 375 field?

A. We will certainly consider that, but we cannot implement it until after removal is complete. Otherwise, current records would get validation errors. But we will move in that direction in the future.

Q. I could've sworn genre headings in WorldCat Discovery records were hot linked, but now they are not in our Discovery. Is this a setting we can change in our Service Configuration?

A. WorldCat Discovery does not display genre headings as clickable links, and we have not historically. While this is not something that can be changed by the user in service configuration, we do support search of the Genre/Form index in WorldCat Discovery as shared in this documentation:
https://help.oclc.org/Librarian_Toolbox/Searching_WorldCat_Indexes/Bibliographic_records/Bibliographic_record_indexes/Bibliographic_record_index_lists/Alphabetical_list_of_available_indexes#G