

2019-10-09 VAOH session

Presentation summary

Patty Treboni, Cynthia Whitacre, Robin Six, and Shanna Griffith presented 1, 2, 3, 4, 5: Everything you wanted to know about stats but were afraid to ask. The main topic was broken down into these sections:

- Retrieving institution or individual library statistics for Connexion transactions by Patty Treboni
- Record Manager statistics by Cynthia Whitacre
- Enrichment & merge stats by Robin Six and Shanna Griffith

URLs mentioned during the presentation:

- AskQC: askqc@oclc.org
- OCLC Statistics: <http://www.stats.oclc.org/cusp/nav>

Member questions

10% merge rate seems low for DDR?

Answer: Actually, this is a high rate. When DDR first started, it merged 2 to 3 percent of records it reviewed. This higher rate can be attributed to improvements in DDR processing and the way we ingest batch records now means DDR bears more of the burden to make sure records match as opposed to the records being matched on their way into WorldCat.

In WMS, can I have statistics for a specific cataloger?

Answer: Not at this time. This is a request we have received before and we do plan on adding in future developments of cataloging reports.

We don't always re-import the bib when we make BIBCO or CONSER replaces, so we can't use our ILS to track that replace-type stat. We manually keep track of these replace transactions (hash marks!) Does anybody have a more automated solution?

Answer: It sounds like that's a function of what you're doing with your local system. The examples given in chat were from a billing report rather than from the statistics portal, so that would be a different way of tracking which was not covered by today's presentation and outside the scope of what we are able to address.

There is a limit to the number of lines in the stats-reports: will that be lifted any time soon?

Answer: With the move to WorldShare Analytics reports the limit to number of lines will be 50,000.

PCC has asked us to report stats on changed/new series authority records and name authority records separately, which I don't think OCLC Stats currently separate out. Has there been any consideration to separating out these NACO changes?

Answer: A further breakdown of the categories has been discussed before. We don't have that in place at this point, but it is something we have been thinking about. If we have the opportunity to revise these statistics in the future it will be considered.

Could you say more about the different statistics categories available under "WorldCat Record Actions"? For example, what is the difference between Update Existing WorldCat Records, Update New WorldCat Records, and WorldCat Replaces?

Answer: Update is about updating holdings, not about updating the record themselves. Update existing WorldCat records would be adding your holdings to existing records in WorldCat. Update new WorldCat records would mean adding a new record to WorldCat and adding your holdings. WorldCat replaces would be using the actual replace transaction to change the master record.

Are there plans to add new statistics to record manager?

Answer: Not at this time. We will be working with the Analytics team in the future to move cataloging reports into WorldShare Analytics but this is not yet scheduled. We will look at improving existing reports in the move and any feedback on cataloging reporting needs is greatly appreciated.

In WorldShare, I only have access to Metadata and Admin. How do I get access to Analytics?

Answer: Libraries can contact support@oclc.org to adjust the roles assigned to librarian's WorldShare accounts to confirm access to reports is provided.

In WorldShare, under Analytics - Reports, why is my only option E-Resources?

Answer: Libraries can contact support@oclc.org to adjust the roles assigned to librarian's WorldShare accounts to confirm access to reports is provided.

Why are there two categories for Produce (Produce Existing WorldCat Records, Produce New WorldCat Records)? I thought Produce was related to getting cards printed and that that product was no longer available. If that's not what Produce means, what does it mean? If that is what it means, why are those categories still there?

Answer: It does relate to producing catalog cards. It is a holdover from the past, we produced catalog card up through 2015. It sometimes takes several years to deal with the administrative part of changing this system. It's one thing to stop sending out catalog cards, it's another to change these categories to get rid of 'produce'. Basically it means the same thing as update, so, produce/update it's the same functionality, it sets holdings in WorldCat for your institution. When you download statistics if you want to combine those categories for your own purposes, you certainly can. The reason for produce existing and produce new, we used to have different billing categories for those back when we had transaction pricing. Now with subscription pricing covering all transactions, there is no longer any distinction other than one is for an existing record and one is for a new record.

Librarian's Toolbox link to OCLC Usage Statistics: https://help.oclc.org/Librarian_Toolbox/OCLC_Usage_Statistics. The entry for *Cataloging* covers the topic *Using OCLC Usage Statistics for Cataloging* which details cataloging reports and cataloging report field definitions.

How are replaced records counted (# of records or by individual replacement actions)?

Answer: Individual replacement actions, so whenever you invoke the replace command.

What automated merge projects have been done on existing records in the last couple of years. (not just recently added records) Perhaps ebooks and eVideos, looking years in different fields, not just in the header. or some other creative area the staff doing manual mergers have seen repeatedly?

Answer: Charlene Morrison outlined an e-serials duplicate project which she and Hayley Moreno presented for the ALCTS Electronic Resource Interest Group at the ALA Annual 2019 ERIG Meeting. They were able to initially identify about 713,000 OCLCE candidate records and merged 157,000 duplicate sets in the first pass. The presentation *Multiplicity amending e serials duplication in WorldCat* can be accessed [here](#) on the ALA Connect website.

We also have a few projects we run on a regular basis with the approval of vendors where we run records through a macro to match on ISBN along with some additional checks. This is usually done on new pre-pub records that normally DDR would not be able to match due to inaccurate pre-publication information.

We have also identified sets of records that we can manually force through DDR when we see patterns we think DDR can now merge.

Is a new record checked for duplication upon creation, or as part of a scheduled scan?

Answer: All records that are added to the database online and through batch load are subject to DDR processing as are records that are replaced online. There is a window of about a week after a record has been created or added before DDR processes it. This allows for changes and updates to the record which may happen right away.

Our library uses Brodart's FLEX account for shelf ready books and MARC records. When they send our holdings to OCLC, sometimes a brief duplicate record is created and our holdings are set on it, instead of being set on a fuller record that already exists for the title. Why does this happen?

Answer: These records come in through our DataSync service. If your holdings were set on a brief record as opposed to a fuller record that might already be present in WorldCat, that means the record sent from Brodart didn't match to the existing record in WorldCat for whatever reason and sometimes the reasons are obscure. The paging may be off a little bit, the date of publication may be off a little bit, if there are enough of those little differences it might not match right away. If you believe this has happened, report that to bibchange@oclc.org and eventually, the records will be manually reviewed and merged if appropriate.

Would it be helpful to email the Connection help contact the OCLC numbers to records that are duplicates?

Answer: Please send any merge requests to bibchange@oclc.org so Metadata Quality staff can merge the records. please see OCLC's Bibliographic Formats and Standards, Chapter 5.5 *Requesting Changes to Records* at <https://www.oclc.org/bibformats/en/quality.html#requestingchangestorecords> for information about reporting change and merge requests to us. There is a section at the bottom of the page that details the different reporting methods available which you may find helpful.

One new headache we are currently facing is where either ISBN 10 or ISBN 13 has the subfield q but the other does not. Is there a process OCLC can run to create the subfield q when it is missing in one?

Answer: Currently, if a record is created online with a field 020 with an ISBN and subfield \$q, the subfield \$q will carry over when the associated 020 field is created. This appears to now only be an issue with DataSync not retaining the subfield \$q along with the rest of the 020 field when transferring the ISBN to records. This is something that would need to be addressed with a DataSync enhancement. It is unknown at this time if the retrospective addition of missing qualifiers could be added to records.

As the NACO coordinator at my institution, I wish there were a way to view all of an institution's authority records for a time period, e.g., ask Connexion to show me all authority records with our NUC symbol in the 040 from 2018. It would be a game-changer for NACO quality control. You can already search for your symbol as "Cataloging Source" in Connexion, but it shows you all NARs you've done since the beginning of time, which on our case gives 40,000 records and makes my head blow up. Maybe there is a clumsy way to combine Cataloging Source and ARN wildcard search?

Even if we could limit to looking for an 040 \$a it would be great and better than what we currently have.

How would you select on last edit date in 005?

Answer: The difficulty with this one is the timeframe as the replaced date isn't tied to your symbol in the 040. With our current indexing, there isn't an easy way to do this type of search. A member has suggested limiting your search by the first four digits in the LCCN, such as In:2018* but that might still give a big group of records. Authority records are not indexed the same as bibliographic records; however, just about everything is indexed in the keyword index in authorities. So you can do some interesting searching in authorities and retrieve records based on words that appear in something like field 670, which you would not really think of as a traditional access point at all for an authority record.

It does not appear the replaced date is indexed in the keyword index. However, the LCCN is indexed in the keyword search, I was able to narrow searches down to the day with this search kw:20150618* as an example.

There was also the suggestion to replace the dlc in these URLs to your institution's MARC Organization Code: <http://id.loc.gov/tools/mynames/dlc/feed/1> and <http://id.loc.gov/tools/mysubjects/dlc/feed/1> and search them with a browser. These are RSS feeds, so if you want more eye-readable results, you will want to use an RSS feed application. The output is displayed chronologically.

I'm finding records for non-youth books with one childrens heading (650 1) when all others are LCSH. Sometimes the childrens heading dups an LCSH and sometimes it's unique. Many are pcc. Any idea why this happens?

Answer: Likely through field transfer, either by the merging of records or via DataSync. When a record comes in and matches an existing record and a particular scheme of subject headings is not present in the record that is already in WorldCat, but is present in the incoming record or the record being merged, that subject heading transfers.

Was there much about metadata at the OCLC Americas Council Library Futures Conference in Phoenix? Any recordings, etc., posted yet?

Answer: None of us were there, so we don't have first-hand knowledge about the content presented or of any recordings which may be posted.

Why doesn't the data ingest program run records through the validation process? It's annoying when you make no changes to a record but it won't validate until you correct some field's data.

Answer: It is the case that incoming records through data ingest are subject to validation. What we do internally is assign an error level to the results of that validation process, some errors are considered far worse than others. Some may affect the structure of the record itself, which is a problem, so we would not necessarily add those kinds of things. But if errors were more superficial, we would go ahead and add that record to WorldCat figuring it is far better to have the library's holdings represented by a bibliographic record even if there is a problem with a tag or subfield code, than to have no record at all. Of course, we try to fix as many of those as we can, but it is easier to fix records that have patterns of errors than records with an off-the-cuff singleton type of error. So this is why you will see records in WorldCat which were added through batchload will not necessarily pass online validation in the same way that records which were entered directly online will pass validation. We are taking data from another library's system and accept some of the errors that are there. Of course, our goal is to fix them up as much as we can, but we also need your help in doing that.

Has the process that moves the order of subject headings around so they are no longer in order as assigned by the cataloger been stopped?

Answer: It is believed this is going to require fixes in multiple locations and this is currently in process.

Might certain pre-pub vendor records be targeted for a merger process, with easier rules for match (i.e. ISBN, fuzzy year, first word or two of title), DDR with specific vendors e.g. brodart

Answer: We are making some enhancements to matching, to match on OCLC control number with some checks in place which should help matching these types of records.

I have started seeing some records with every year the book was published in the 264 4. I had a third edition with 2019, 2010, 2002. Is this going to be a new practice??

Answer: Hopefully not. Please send those to bibchange@oclc.org so we can investigate.

2019-10-23 VAOH session

Presentation summary

Patty Treboni, Cynthia Whitacre, Robin Six, and Shanna Griffith presented 1, 2, 3, 4, 5: Everything you wanted to know about stats but were afraid to ask. The main topic was broken down into these sections:

- Retrieving institution or individual library statistics for Connexion transactions by Patty Treboni
- Record Manager statistics by Cynthia Whitacre
- Enrichment & merge stats by Robin Six and Shanna Griffith

URLs mentioned during the presentation:

- AskQC: askqc@oclc.org
- OCLC Statistics: <http://www.stats.oclc.org/cusp/nav>

Member questions

Does this work with Google docs? We no longer have access to MS Excel.

As long as you are able to open or import Excel files into Google Docs, this should work. On the OCLC Usage Statistics page, you can just highlight and copy the information and then paste it into whatever application you are using.

I tried to open the slides from the Oct. 9 presentation but was unable to. What format are they in?

The AskQC presentations are PowerPoint slides. An alternative to Microsoft Office or Google Docs might be to try Libre Office. You may need to speak with someone at your institution who sets up your office applications.

Is there an overlap of the type of reports available on WorldShare and Connexion?

There are very similar statistics but they are not kept in both places. The WorldShare statistics are only available in WorldShare, the Connexion statistics are only available through the OCLC Usage Statistics webpage. While the reports are not exactly the same, they are very similar.

We're a merge library (and we love it a lot!) but our PCC stats don't get recorded on our merge authorizations. Is there a workaround other than doing work in 2 different authorizations (like doing the merge in one and then the authentication in another?)

Yes, you would need to use two different authorizations for that.

Does Record Manager report for a separate set of accounts, or can one get the same report for an account from either Connexion or Record Manager?

The Record Manager reports as shown in the presentation under the Analytics tab reports all the transactions done using the Record Manager interface. It does not include transactions done using the Connexion client or

browser interfaces. Another way to say it would be if one person worked in both Record Manager and Connexion and wanted to get a total of their work for the month, they would need to go to both places and combine the numbers.

If you have not used Record Manager, the Connexion authorization numbers do not work in Record Manager. It is an entirely different method of signing on and requires a different set of authorizations.

I listened to the 9 October AskQC session. To confirm .. the update and produce function are basically the same for counting purposes?

Yes.

In Connexion statistics, when you want to receive reports through e-mail, and you want it sent to more than one address, do you add a semicolon between addresses in the dialog box? Or does each address need to be added separately?

Multiple addresses can be entered in the E-mail Address dialog box. It appears each address can be separated by a semi-colon, comma, or space.

We are a merge library, and I am wondering if it is possible to obtain the merge statistics for our institution? (Unless you already covered this and I missed it?) Thank you.

Yes. if you are a Member Merge library and would like to receive your merge statistics, contact your reviewer or AskQC@oclc.org and we will be glad to get you set up with that.

I have a question regarding export. This morning, I tried to submit NACO statistics for midyear report. I got the statistical data. How does the export mechanism work? Someone told me to enter them manually and was given a weblink. But I could not login as my login info was invalid. Whom shall I email? Where can I get more info about NACO participants submitting statistics?

You will need to take that question to your contact or the NACO contact at the Library of Congress because they run the statistical reporting mechanism for where you enter those statistics. This is totally separate from OCLC and we don't have any control over that. You can get statistics from OCLC, but as for manually entering them, that is a Library of Congress system where you are reporting them.

The general NACO email address at the Library of Congress is naco@loc.gov.

Also, this has been an active topic on the PCC participants list, that is another place you can go for more information.

For NACO questions you can contact Luanne Goodson at authfile@oclc.org.

Do we need to contact NACO to set up our stats with OCLC?

No. The earlier question was about reporting NACO statistics to LC. You can get your statistics for what you do using OCLC interfaces, as described in the presentation today.

Does OCLC track instances of controlling headings?

Sort of. We don't keep statistics on controlling headings. We do track headings that get changed which have already been controlled and then if the authority record changes, we get notices of what has changed through the automated updating process.

How do you get to be a library that can merge records?

If you would like to participate in the Member Merge Program, email AskQC@oclc.org. Laura Ramsey will respond with more details.

Currently, OCLC is harvesting our collections in our DSpace repository. Now we have created a new collection. How can I get this harvested too?

While a great question, this is out of our area knowledge and expertise. The best thing to do would be to contact OCLC support or email them at support@oclc.org.