

2019-06-12 VAOH session

Presentation summary

Shanna Griffith explains how duplicate records can be added to WorldCat just by a difference in the choice of title by the cataloger, even though all other information between the records is the same. Laura Ramsey provides an update on the Member Merge Project, how your institution can get involved, and the impact it has made in the ongoing quest to eliminate duplicate records in WorldCat.

URLs mentioned during the presentation:

AskQC: askqc@oclc.org

Bibchange: bibchange@oclc.org

Member questions

Topic Specific:

Query - the LCSH heading that was carried over had Fiction not Juvenile fiction. This means that I would have to make a correction before accepting this record.

Answer: Headings that contain incorrect information, such as subfields, can be corrected either prior to the merge or after the merge is complete. Metadata Quality staff do try to clean-up the records using a clean-up macro and make any necessary corrections that they find as they are merging duplicates.

How does the clean-up macro work? Does it look at the codes and change the fields, or look at the fields and change the codes?

Answer: It is a little more complicated than that. The internal macro that we use has more than 30,000 lines of code, so it is very complicated. It looks at many, many factors trying to clean up obvious issues and problems in the record. Internally, we call this our QC or Quality Control macro. It works just like any other macro, it's just longer and more complicated.

How many members are merging serial records?

Answer: At this time the University of Oregon and UC San Diego are the only institutions currently merging serial records, with another institution in training.

My notes from an AskQC session about DDR say that the 300 field could have two subfield \$a's such as 1 score ; \$c 28 cm + \$a 4 parts. Is it okay to repeat subfield \$a in the 300 field

or should it be subfield \$e? At the time, DDR (Duplicate Detection and Resolution) was ignoring subfield \$e and this caused records with CD + book to be merged incorrectly.

Answer: DDR does take into account accompanying material, however that information should be recorded in subfield \$e of the 300 field, not subfield \$a. If a record has accompanying material recorded in 300 subfield \$e, it will not get merged into a record that does not have accompanying material. At one time, DDR used to handle this differently. We changed this a few years ago so that accompanying material is taken into account in the Duplicate Detection and Resolution software merging process.

Since this is for PCC only, is there a path for those of us who are not to request a merge from a participant, rather than going through bibchange?

Answer: No, there is not. We still ask that you report duplicates to bibchange, if you are not a Member Merge Project participant.

Does NACO funnel participation count as PCC participation?

Answer: Yes, if your institution is a NACO funnel participant, you may contact us to participate in the Member Merge Project.

With this new program, how long on average does it take to process a merge request?

Answer: We don't track how long it takes to resolve an individual request. We have requests that come in for different formats that are addressed by different staff members, so the turn-around time can vary for different formats. We have some formats that are up-to-date and we have others that have several months of backlog.

Is merging happening more quickly now that some members can merge as well?

Answer: The majority of the members who are merging are not merging from the backlog here at OCLC, but are merging things that they come across in their daily workflow. While it may be making things faster because they are coming across less duplicates as a result, it is not diminishing the backlog that has been submitted to OCLC, at least not at this time.

How long does it take to merge records when I send a merge suggestion to bibchange@oclc.org?

Answer: That does depend on the format. We do have different turnaround times depending on the format and the number of staff available to review those requests. The books format backlog is our most extensive backlog at close to a year or more at this point. In general, without breaking it out by format, our backlog consists of approximately 40,000 requests to give a sense of how much we have to work on.

We use RecordManger, if we wanted to participate, would you have protocols, scripts, or tools for RecordManager or is Connexion used?

Answer: At this time, it is a Connexion only function. We don't have it built into RecordManager.

How do we know when to report merging records when one has been accustomed to seeing so many records for a particular manifestation (incomplete records, half-finished records, etc.)?

Answer: If you believe that they are duplicates when you look at them, feel free to report them. We can make the determination when reviewing them if they are indeed duplicates and merge them, or if not and leave them as separate records.

My institution has done a reclamation project which created a lot of duplicates in OCLC. Could this have been prevented on our end or yours? Most of our records had old OCLC numbers on them.

Answer: We would have to see some records and do some investigation in order to figure out what would have helped out in that situation.

Does field 035 (System Control Number) with subfield \$z have something to do with merging?

Answer: No, not that we are aware of.

Is there a quota for how many records an institution should merge per year?

Answer: No, we did not set a quota. It is really dependent on the institution and what they are able to fit into their workflow.

If I report a record that needs to be merged, do all member libraries get that message, or only OCLC?

Answer: Only OCLC get merge requests at this time.

Do you pass the requests submitted on the web to the libraries in the project, since you have 40K+ requests?

Answer: We would love to, but we have not had that request from any of our participating institutions. When first starting in the project, some institutions don't have duplicates put back that they've come across in their daily workflow, so we will use requests from the backlog for training purposes. Once the institution becomes independent, their time spent merging is for those that they come across in their day to day work.

I rarely find duplicates for the books I catalog, but I find quite a few when I catalog scores. How would this play out for training purposes?

Answer: There is training documentation for merging scores duplicate records. So, if you only work with scores material, you may request to become a participant in the project to get training to merge only scores duplicates. Initially when we started this project, we had the requirement that all institutions went through the books format training. We then realized with other institutions requesting to join that didn't catalog books, it didn't make sense to have that requirement. So we are working with a few institutions that are training on the specific format they are interested in.

When we merge records, we have to transfer Google links. Two questions about this related to serial merges, which can have several hundred Google links to move. 1. Is there a macro that can help this transfer? 2. Is there any thought going into not moving all these links in the future?

Answer: When you do a merge, those will automatically transfer during the merge process. We are in discussions with changing the Google process, and all of those Google links is something that we have talked about improving.

Do Google links transfer automatically into CONSER records?

Answer: Yes, they do.

I have a list of a large number of records (about 80) that I believe were improperly merged. They each have an 019 field and inconsistencies in the formats across the fixed fields, 006/007, and 33x fields. What would be the best way to send such a large number of these at once? I've been simply typing up the details in a Word document, is that something that you could work from?

Answer: Yes. Record numbers for those records that you believe to have been incorrectly merged can be put in either a text file or directly in the email message and submit them to bibchange. We will review them on an individual basis to determine what happened and recover, if necessary.

I remember the 1st cohort required a massive time commitment from OCLC staff, so much so that expanding the project was in serious jeopardy. Now, 14 more institutions have been added, and other PCC institutions are encouraged to join the project. What changed to make it possible for so many more institutions to participate?

Answer: Several things probably played into that - a learning curve, being one. We were starting up something new with the 1st cohort and we really didn't know what to expect from that. So we have continued, over the years, to improve and tweak the program. We also almost doubled our number of staff that are reviewing, so we were able to take on more institutions. Using an external review also helped with that process. We realize that this is really important to our member. It does seem like a big time commitment for OCLC but in the end, it really does pay off. As you saw from the number of records that the members are merging, in the end it's a win-win.

Do you have numbers of members participating for the different formats (for example, scores) and the number of score records merged?

Answer: We get the OCNs for all the merges that are done, but we don't have any way to determine what the format is for that. Those participating in the project can go to the OCLC Usage Statistics and see the numbers that they have merged, but that is all the information that we have available from our stats.

If I send a merge suggestion to bibchange@oclc.org, will I get any feedback about what happened? When I send a suggestion, if they don't get merged, I would like to know why.

Answer: With the amount of duplicate requests that we receive in a day (over 100 requests), there is no way to provide feedback with the result of the request while trying to process as many requests that we possibly can. We are in the process of streamlining merge requests to process them quicker. This process strips out the contact info to harvest the OCNs reported, leaving no way to provide feedback afterwards. For reports of believed incorrect merges, we do provide feedback as to whether the merge was actually correct because the merged record contained incorrect fields that transferred during the merge process, or that the merge was in fact incorrect and the record(s) has been recovered.

I've been working on matching our old WLN records to the OCLC version, and I've found a number of records that, based on the variety of notes (5xx) are mergers of special editions (e.g., artistic illustrated editions, etc.) that would seem to have been better on their own record.

Answer: If you believe that records have been incorrectly merged, please report those to bibchange so that they can be investigated and pulled apart if necessary.

Does field 019 have something to do with merging records? Sometimes I take the OCLC 035 field and search it in OCLC and it appears in the 019 field and not on top where the OCLC number is and is really confusing.

Answer: The 019 field contains the OCNs of the records that have been merged into that record. The OCLC number that is found in your the 035 field of your institution's record could appear in the 019 field of the master record in WorldCat, showing that record has been merged with the master record in WorldCat. The reason for retaining these OCNs in the 019 field is so that if your institution does have the OCLC number for one of those and search on it, it is still indexed and you can find the record that it has been merged to. So it is very important to retain these so that the library can find the remaining record.

Besides the 019 field, are there additional tags to identify merged records?

Answer: No. The only way to identify a record has been merged is by the presence of that records OCN in the 019 field.

Are the 33x fields taken into account when merging?

Answer: Yes. The 33x fields contribute to the formulation of the material type and are taken into consideration for DDR (Duplicate Detection and Resolution) and other automated processes. For manual merging, they are given less consideration because we see the material type intended in the record and can determine if the 33x fields are correct or not.

Can participating institutions merge across formats? So if one record is for a CD plus text and another for the same item is for a text plus CD?

Answer: Yes, these could be merged, but there could also be a good reason to leave them as separate records in WorldCat. It depends on the context, and you would want to consider each case individually. If one was poorly cataloged and the other is much better cataloged and has a lot of holdings on it, that might be a reason to merge them. If both are being used equally and the cataloging institution had a legitimate reason to place one as the predominate material over the other and some other institution made the opposite decision, those could continue to be separate in WorldCat. Bibliographic Formats and Standards chapter 4, When to Input a New Record states that if you do encounter a record where you disagree with the choice of predominate material, to use the record rather than creating another one.

Where do I find the number of records I have merged within OCLC Usage Statistics?

Answer: When you log into the Usage Statistics, it will be the second-to-last category at the bottom that says Merged Records. When you open that, you will see Institution Monthly Merged Records. Within that, you can enter your institution's symbol to retrieve the records merged for the given range. Please reach out if you are having trouble accessing your stats.

Are you able to merge cataloging records with non-English language of cataloging?

Answer: Records with the same language of cataloging (i.e., both records are German, English, Spanish, etc. language of cataloging) can be considered candidates for merging if their descriptions match for the resource described.

What if the language of cataloging seems to have been in error?

Answer: We ask that you report those to OCLC for further investigation. We can check the history of the record to see how it came in to us and if it was possibly incorrectly changed by another institution. The revision of Bibliographic Formats and Standards chapter 2 will include an updated section for language of cataloging records and how to handle hybrid records.

General questions:

One of the problems I've seen with using Data Sync to merge records is that libraries can do what they want locally with subject headings, etc. and I see a lot of bad headings coming in that won't pass OCLC's validation process.

Answer: So, this is a comment about OCLC's Data Sync loading. This is not about Member Merge or people manually making decisions about what to merge. This is about how our matching software, in the loading process, matches the records and there are fields that are transferred from the incoming record to the existing record in WorldCat. We have just recently made some adjustments to that field transfer process, that hopefully will improve things so that not as many things that are perceived as bad headings will be merged.

We've come across records for online materials with no 856 fields. Under what circumstances are these made and can we add 856 fields to these records?

Answer: Yes, please add 856 fields to those records. Sometimes we are getting records from vendors, where they don't have a generic 856 field to supply because that vendor generates only individual ones for an institution. Ideally what we want in Worldcat, is a much more generic 856 that anyone could click on to at least get to a sign-on page for the provider.

Do you know how many enhance/replace transactions by a member library are done and what years of publication generally are included?

Answer: We do have statistics on enhancements and replace by member libraries, but we do not have it broken down by years of publication.

Does OCLC create a 050 field from a 090 field during quality control editing?

Answer: Yes, the QC macro that we run on records does change the 090 to 050, if not already present in the record.

How would you be able to tell if a record was derived from another.

Answer: There are various areas within a record that could be a clue that the record was derived from another record.

Couldn't the derived record be in error? Creating a new record when it was not justified?

Answer: It depends on the records. You have to look at the metadata, such as if things are coded incorrectly or if there is cataloger supplied information such as an edition statement to help indicate that it was intended to be a unique record. If the records do not have that differentiating information, then it will be hard for a program to make that judgement, let alone a human, to make the determination that they are in fact duplicates or are meant to represent different resources.

Why does OCLC allow subject headings and genre terms in multiple languages within one record when the language of description is only for one language? In our library consortium, we delete all non-English subject headings, and we have to spend effort removing non-English after importing (particularly for the large number of French and German headings in those records).

Answer: The language of cataloging code found in the 040 subfield \$b is specifically about the language of the description, as correctly stated in the question. So, that doesn't say anything about what kind of subject headings that can be on the record. OCLC allow these because we are membership organization, and there are member libraries that are out there that may want French language subject headings on records that they catalog in English, or may want Spanish language subject headings for records they catalog in English, or English language subject headings for records they catalog in Spanish. We don't want to impose a policy on member libraries that would preclude that. That is why you will see a huge mixture of subject headings on records of all types of languages of cataloging.

When we are upgrading a CIP record, why are we unable to change the encoding level from pre-publication (8) to full-level?

Answer: There are multiple reasons. To change the encoding level, an institution would need to be able to create authority records if they are adding subject headings or access points, particularly name access points to the record. Those records, especially those from LC, are PCC records and need authority records to back up the heading. PCC libraries can do that upgrade, and change the encoding level. We also have some CIP upgrade program participants that are specially authorized to do that.

I've compiled a list of institutions that add institution specific URLs to 856 (ezproxy addresses). I think they are editing them through WMS. Can I send this to you? I recently replaced a lot of old records with newer ones with a WorldCat query. It added a lot of bad links to our catalog.

Answer: Yes. One of the things that we do when we are working on records for various reasons, is try to delete local URLs when noticed. We realize that those clutter up WorldCat, so if you have noticed a pattern, please send those along to us.