

## 2021-06 VAOH Session

### Presentation summary

Laura Ramsey and Shanna Griffith presented: *Evolution of a WorldCat record*. This presentation covered how many OCLC processes work, such as matching, merging, and field transfer, just to name a few, and how this information will be presented in the new comprehensive documentation that will serve as a resource to provide answers to processing questions that we get. This documentation will be publicly available later this year in a centralized location on the OCLC help site.

### Resources and URLs mentioned during the presentation:

Cataloging and Metadata Community

Center: [https://www.oclc.org/content/community/en\\_us/cataloging-metadata.html](https://www.oclc.org/content/community/en_us/cataloging-metadata.html)

Bibliographic Formats and Standards (BFAS) <https://www.oclc.org/bibformats/en/quality.html>

To subscribe to any OCLC distribution and discussion

lists: <https://www.oclc.org/content/forms/worldwide/en/internet-subscription.html>

Searching WorldCat Indexes for field

655 [https://help.oclc.org/Librarian\\_Toolbox/Searching\\_WorldCat\\_Indexes/Bibliographic\\_records/Bibliographic\\_record\\_fields\\_and\\_subfields/6xx\\_fields](https://help.oclc.org/Librarian_Toolbox/Searching_WorldCat_Indexes/Bibliographic_records/Bibliographic_record_fields_and_subfields/6xx_fields)

Reports available for Data Sync processing: <http://oclc.org/datasync-reports>

### Member questions (2021-06-08)

**Occasionally I find monograph records (ranging from sparse to pretty complete) in LC's catalog, but there is no record at all for the work in WorldCat yet. What is the workflow for DLC records to enter WorldCat and is there a particular action or record change at LC's end that triggers the export of that record to WorldCat?**

The Library of Congress (LC) distributes most of the records they catalog and we receive weekly distributions from them; however, they do not distribute everything in their catalogs. If you are finding a record in LC's catalog, that does not necessarily mean we have received it, so if you are wanting to know about a specific record, please send a question to [bibchange@oclc.org](mailto:bibchange@oclc.org).

**Can we have a link to this documentation? Because I would prefer to read it.**

We are hoping to have the documentation finished and posted later this summer to the OCLC Community Center and the OCLC Help website.

**Is language of cataloging a factor in this processing? Are there any differences in the process on that basis?**

The language of cataloging is taken into consideration in DataSync processing and deduping WorldCat. For example, we want to match an English-language cataloged record to an English-language cataloged record and a German-language cataloged record to a German-language cataloged record. But not match

an English-cataloged record to a German-cataloged record. Beyond that, there really is no difference in the processing, we compare the various elements of the records, in the same way, regardless of the language of cataloging. Taking field 300 as an example, we're normally looking at things like the coding of the record to determine format and then the extent in terms of the numbers that are present, rather than actually comparing the terms, as different terms are used in different languages to indicate the same thing.

**DataSync: from my understanding of it only "record matching" is available at this time. Will "number matching" be available again at some point in the future?**

Right now, we are in the process of doing some testing to incorporate number matching back into DataSync matching. We do not have a definite date or timeline when that will be installed, but it may be as soon as in the next month or so.

**Are the OCLC numbers in the 019 Field entered chronologically? I wonder how they are ordered in the one Field.**

As records are merged the OCLC record numbers are added to the 019 field in numerical order, regardless of when a particular merge took place in relation to any other OCLC record numbers in the 019.

**Does this documentation revolve around Data Sync? (Trying to figure out the context.) Does the same process describe also apply when vendors submit e-resource records to Collection Manager?**

The documentation will cover DataSync as well as many other processes. For example, Collection Manager and Duplicate Detection and Resolution (DDR) are two of the wide array of other processes that will be covered in the new documentation.

**Do you have any rough estimate on the percentage or proportion of WorldCat records that are added manually via Connexion or Record Manager versus through a batch or other process?**

This comparison between manual and automated processes to add new records is not something that is currently tracked; however, the vast majority of new records added are contributed by automated batch processes.

**Does the process recognize hybrid records, i.e., records with evidence of different language-of-cataloging practices? How is merging done in that case?**

The process looks at the 040 \$b for the language of cataloging code only in determining the language of cataloging of a record. If a record were a hybrid where the 040 \$b code did not match the language of cataloging in the description of the record, then there would be the possibility that the record could be merged into a record with the same language of cataloging code and fields in the different language could end up getting transferred.

**Is there any move to make the matching algorithms between DataSync and DDR more uniform? Frequently DataSync will input my records as new (even if they contain OCLC numbers), but then these new records are merged later on with pre-existing records via DDR.**

With the updates coming to the DataSync matching that was mentioned earlier, the match rates for records coming in via DataSync should improve. This is not necessarily to make DataSync and DDR matching algorithms more uniform, but it will improve the match rates in DataSync processing.

**Is every number in a 019 field represented by an institution code in 040 \$d?**

Not necessarily, if no information was transferred during a merge, then the symbols present in the 040 will not transfer to the retained record's 040, only the OCLC record number will be added to the 019 field. If data does transfer during the merge, then the symbols will be added to the 040 unless already present.

**I see a number of PCC records that use the \$v Juvenile ... with annotated children's headings, which is against the guidelines for annotated children's headings. Is there anything I can do to these other than reporting them as an error?**

If you are not a PCC library, so that you're not able to change those headings to their correct form on your own, I would say, go ahead and report them and also note that there are others in the database. We have been doing a little bit of work with juvenile subject headings of late and noticing cases where we have juvenile subject headings that include subdivisions, juvenile fiction or juvenile literature that should not have those subdivisions, as well as LCSH headings as well that are related and need to be adjusted, so go ahead and report those errors.

**MARC doesn't have a method of recording data provenance, does OCLC keep this information internally, e.g., who input a certain field?**

No, we don't because the MARC format doesn't have that we do not have that same kind of information internally. So the only record of a change to an existing bibliographic record would be the addition of a symbol to the 040 field. But once you have a lengthy 040 field with many \$d's, and a lot of fields that have been added or changed in the record, there's no way to match up who did what. We do have a history of the record so we can look back at certain points in time and compare before images and after images to tell what happened, but not in the sense that the question was asked in terms of, can I tell by looking at a field who added it for example.

**Is there a way to see the way a record looked prior to the most recent change?**

We do have an internal history tool that will allow us to look at records prior to a change transaction as long as it happened after April 2012.

**I've seen a number of records for media lately that put personal name \$a and \$d information in separate sets of 700 fields. How would the processing handle such cases? Are problematic batch loads ever excluded from routine processing?**

Unfortunately, I don't think that we have anything in our data prep information processing that I explained. We do a lot of clean-up, but I don't think that's something that our processing can handle. Definitely bring these to our attention so we can work with the database specialists for that particular project. We can communicate back to the institution, so they can fix future records and there is a possibility fields could be excluded from future processing if it's something they are not able to fix.

**Would the number of 700s be a factor in choosing which record to retain?**

I can say that it does figure in to some decisions in part. The record retention hierarchy that we have in place starts off by looking at things like codes and field 042 and the source of the record such as the Library of Congress versus an OCLC member. But once 2 records are in the same rank, then we consider the number of fields that are present and the number of holdings in combination and so a record that appears to be more complete because it has more access points and field 700, could possibly win out in that comparison.

**How much of what you just described is automated?**

All of it.

**I work in a library that uses multiple OCLC symbols. If a record was created using the wrong authorization, so the wrong symbol, is there a way to correct this? Should we report these to OCLC?**

Yes that something that we can correct for you. If you send a message to [bibchange@oclc.org](mailto:bibchange@oclc.org), with the affected record numbers if you have them, we can make the change for you if you would prefer that a particular symbol as associated with a record. That's something that we have to do, it's not something that our users are able to change in a record.

Member questions (2021-06-17)

**Will the documentation give a better idea of the various OCLC codes used in the 040 field?**

No, we are not including that; however, you should be able to find information on the codes we use and the purpose of those codes in Bibliographic Formats and Standards (BFAS) at <https://www.oclc.org/bibformats/en/quality.html>. This might be good information to include, so I'll bring that back to the team that's working on the documentation.

**I've seen many fields transferred into merged records with \$5 subfields. It would be great if you could remove these fields or not transfer them.**

If you could send a message to [bibchange@oclc.org](mailto:bibchange@oclc.org) with some examples, we're happy to look into it and see if those fields are appropriate for the record or not. Then if they're not, of course, we will be happy to remove them.

**Are very brief bib records from Book vendors still being loaded?**

We have what we refer to as sparse record checks, so we do have criteria that records need to meet in order to be added to WorldCat. So, the records that you refer to are obviously meeting those the criteria. I don't know if you have any specific examples that you're asking about, again, we're happy to take a look to see if there are some issues, but we do try to set criteria to make sure that we're getting records that have enough metadata so they can be searched, retrieved and be discoverable.

**When I am editing an OCLC record via Connexion Client, am I directly editing the WorldCat record that appears on [WorldCat.org](http://WorldCat.org), or is there some distance between the records as they appear in those two different places? Would you see that change more or less in real-time, or is there an amount of time between that edit and its appearing in [worldcat.org](http://worldcat.org)?**

You should be able to see the edits you make using Connexion client in the [WorldCat.org](http://WorldCat.org) display of the record after replacing the record then refreshing the webpage. Additionally, there can be information from external sources displayed in the [WorldCat.org](http://WorldCat.org) record display that does not exist in the WorldCat bibliographic record.

**When the new Connexion client is going to be deployed?**

We don't have an exact timeframe yet. The testing is happening right now with the field test. I would predict in the next few months, but I don't think we have a specific timeframe yet.

**If you do a title search for "untitled" and limit to 2020 on, there are hundreds of brief records that appear to be from book vendors- how long do these stay in OCLC?**

That's actually one of those things that's on our list to clean up periodically. So, it looks like we need to put that on our list to clean up sooner rather than later. We do try to take care of those and either merge them into duplicate records or delete them.

**I have a totally unrelated question about indexing: Is it possible to "turn on" indexing for MARC field 655\_4 in Connexion or Record Manager so that field can be used for keyword searching?**

If our [Searching WorldCat Indexes](#) document is correct and if I'm reading it correctly, the 655 is indexed in the subject [su] index and in the keyword [kw] index, in addition to the specific indexes that take into consideration the 2nd indicator. So a 655\_4 is indexed and should be searchable in both the subject and keyword indexes.

**What is the usual time from adding a record, to time matching, replacing, determining if keeping a record, a decision is made?**

We have several different flows that occur with our DDR duplicate detection program so, how a record is added to WorldCat affects when it will go through DDR processing. For example, records added via DataSync generally get into the DDR queue within several days and processed within 48 hours. For a record added or modified online, it takes 7 days to enter the DDR queue.

**For the last week or so when I login into OCLC the message is very brief and says: Welcome to the OCLC Connexion™ service. You will be using the service in Master mode, and then there are phone numbers listed. Is there something going on with OCLC Connexion right now?**

The message of the day was decommissioned in March so a message of the day is no longer being posted. To elaborate, this is related to the new client version, which is in field-testing right now, where there will no longer be a message of the day.

**How different will the new Connexion client be from the current program?**

Except for the opening screen, which looks a little bit different, it looks very much the same as it always has. The changes are largely, although not exclusively underneath the hood, if you will, behind the scenes. The idea was to upgrade the client to be better compatible with more modern and up to date Windows software and so on. So, it won't look a lot different, but in theory, it will be behaving better. Additionally, some things such as glimmer clustering and institution records that are talked about in various places in the connection client, those references have been removed. If I remember correctly, the help system has been moved to the OCLC website. So when you access help, you'll be going to the OCLC website rather than within Connexion itself.

**They are not currently indexed, and we would love them to be. Where could a request be made?**

- **Random feature request: Add a keyboard shortcut for setting holdings in Record Manager. Thank you all for what you do!**
- **It would be very useful to us to have the 655\_4 keyword searchable under "genre"**
- **We'd like to be able to limit to a local genre string, and this doesn't work. One cannot browse by genre Theses -- Biology and one cannot do a genre keyword search for Theses Biology. If you do a subject kw search on those words, you'd get too many extraneous records**
- **I should say the strings are subdivided: 655\_4 Theses \$x Biology. The browse ignores subdivisions**

Record Manager enhancement requests can be made through the [Cataloging & Metadata Community Center](#) or send a message to [askqc@oclc.org](mailto:askqc@oclc.org)

### **How long does it take a NAR to get into VIAF?**

Data from the LC/NACO authority file is sent each week and the VIAF database is updated weekly as well. Depending if a change or addition was made early in the week, you might see it in VIAF the same week, but if you did it later in the week the change might not be reflected in the VIAF database until the following week.

**I did not realize until more recently that 505 field is an optional/optional entry standard. And I wondered why. The 505 field is very important for resources like song albums and collections of children's tales, anthologies, etc.**

The the input standards of mandatory and optional and so on were originally set long ago by an advisory group of catalogers from member libraries and were partially based on the Library of Congress national-level demographic record. Fields such as the 505 contents note, of course, are appropriate in many cases and are extremely useful as you point out, such as for sound recording song albums, collections of stories, anthologies, and so on. But the 505 is also not found in many things, such as a fictional novel, for example. Input standards aren't meant to indicate if a field is worthwhile or not worthwhile, they are meant to be kind of floors, not ceilings, so if any optional field is appropriate and useful in a particular instance, then yes, include them.

### **Will there be new training on the Connexion Client?**

We should have some info on this soon — stay tuned!

### **Remind us what GLIMR is?**

GLIMR was an experiment that OCLC undertook some years ago to cluster bibliographic records together. It evolved into the clustering you now see in WorldCat.org.

### **Follow-up on client: will we still have access to our current text strings and macros?**

Some of us did do some testing prior to the field test and all of our personal macros that we use were made available. I don't use text strings, but I believe both the text strings and macros will transfer over to the new version.

### **VIAF clustering is so weird sometimes and they cluster things or not cluster for same authority records**

Sometimes it gets it right, sometimes it doesn't. VIAF clustering is determined by algorithms, which are not static, they are always being worked on and refined. So, if you see any VIAF clusters that are incorrect, please report them to [bibchange@oclc.org](mailto:bibchange@oclc.org). We manually edit the clusters as needed and can work with the VIAF team to see if there's a bigger underlying issue when indicated.

### **When the new client is available, is it an update or a whole new download?**

It will be a completely new download.

**I had seen a lot of clusters that are not "right" and reported it but months had passed and nothing is been done**

We get many requests for editing the VIAF clusters, so we apologize for not getting those completed as quickly as we would want to. But, definitely keep reporting them, we try to work on them as much as we can.

**How is the consolidation and streamlining of Encoding levels going? I have been seeing level 1 records and I thought those were for old LC copy that was input without much checking of the book.**

We're just working on the conversion of encoding level K records, that's the first encoding level we decided to try to take care of. I just checked today, we still have about 16 million records to go. So we have a little while yet to work on those. I think we were talking about moving on to encoding level M, but that's a huge number and we might break that down somehow or another, but I think that may be next.

**I have also had that experience, Ivan. Are there plans to open up editing privileges for VIAF?**

At this point in time, I do not believe there's any discussion on that being developed for VIAF.

**Will vendor records continue to be coded M and not at what level they were input as?**

We actually are still having discussions on how we're going to evaluate incoming records and how to code them. We haven't decided yet so unfortunately I have nothing more concrete to share.

**Are vendor partners are allowed to use WorldShare collection manager?**

Yes, if they pay for a cataloging subscription.

**What does "VI" in VIAF stand for?**

VIAF stands for Virtual International Authority File.