

2018-11-14 VAOH session

Presentation summary

Jay Weitz presented on the topic of the OCLC MARC update process. MARC21 has evolved over the last 50 years, and Jay gave a history of MARC21 and the field update process. He talked about the organizations responsible for approving new fields, the MARC Advisory Committee (MAC), including MARBI (1973-2013) and the MARC Advisory Committee (2013-). He shared the list of current MAC members, both national libraries and other constituent members and how they work together.

Using the MARC field 647 as an example, Jay walked us through the process OCLC followed in getting this field defined for use for FAST headings.

The MARC Update Timeline:

Idea -> Discussion Paper -> Proposal -> MARC Update ->OCLC-MARC Update

The MARC Code Timeline:

MARC Code Requested -> LC Technical Notice -> WorldCat Validation Install

For your reference, URLs mentioned during the presentation

- MARC Advisory Committee (MAC): <http://www.loc.gov/marc/mac/index.html>
- MAC Terms of Reference: http://www.loc.gov/marc/mac/MAC_ToR.html
- Faceted Application of Subject Terminology (FAST):
<https://www.oclc.org/research/themes/data-science/fast.html>
- Historical links to MARC Discussion Papers from 1995 to the present:
<http://www.loc.gov/marc/mac/list-dp.html>
- Historical links to MARC Proposals from 1995 to the present: <http://www.loc.gov/marc/mac/list-p.html>
- Historical links to all of the MARC 21 Format Updates from 2000 to the present:
<http://www.loc.gov/marc/status.html>
- OCLC WorldCat Validation Release Notes: <http://oclc.org/oclc/validation-release-notes>
- OCLC Bibliographic Formats and Standards: <http://oclc.org/oclc/bfas>
- OCLC-MARC Local Holdings Format and Standards: <http://oclc.org/oclc/lhr>
- Searching WorldCat Indexes: <http://oclc.org/oclc/indexes>
- Authorities: Format and Indexes: <http://oclc.org/oclc/authorities>
- LC Guidelines Supplement to the MARC 21 Format for Authority Data:
<https://www.loc.gov/catdir/cpsol/lcmarc-suppl.pdf>
- Genre Terms for Makerspace Materials:
<https://digital.library.unt.edu/ark:/67531/metadc1132759>
- LC Technical Notice (May 25, 2018): <http://www.loc.gov/marc/relators/tn180525src.html>

Member questions

Yesterday I noticed that a particular organization code (DOT) is obsolete and has been supplanted by ONET. Should the DOT be removed? It makes a difference-- Creative writers is in DOT but not in ONET.

Answer: If a code has been made obsolete by the Library of Congress, it should have a dash in front of it. If it doesn't have a dash in front of it, it is still a valid code and can be used. If the thesaurus or reference document the code corresponds to has been superseded by a later edition or a more up-to-date list but is still a valid MARC code, you can continue to use the MARC code in the field.

<https://www.loc.gov/standards/sourcelist/occupation.html> still has the code dot listed for the Dictionary of occupational titles.

There have been numerous duplicate DLC records being added. Some have been reported and quickly deleted but there is still a lot that DDR seems not to pick e.g. 1061860203 and 1043516446 both have DLC in 040\$a. Why are these being added?

Answer: In general, DDR can't catch everything, and we have designed it to be extremely careful and to err on the side of leaving a duplicate rather than incorrectly merging records that shouldn't be merged. If you have record numbers that DDR has missed, do report them to bibchange@oclc.org. It may take us some time to get to these, but it does help us to get these reports as they can help us find other patterns of issues that we can address.

For the specific records reported, #1043516446 is a Library of Congress contributed record as indicated by the symbol DLC in the 040 subfield \$c. #1061860203 is a member contributed record as indicated by the institution symbol TXN in the 040 subfield \$c. This record was contributed on November 8, 2018, which was only 6 days ago. For records contributed through Connexion and Record Manager, there is a 10-day grace period before DDR evaluates them as potential duplicate records.

Field 040 subfield \$a is used to record the original cataloging agency, not the transcribing agency which is recorded in subfield \$c. For more information about field 040 and what the different subfields are used for, please see OCLC's Bibliographic Formats and Standards, 040 Cataloging Source at <https://www.oclc.org/bibformats/en/0xx/040.html>

Where should we report issues with FAST headings? I recently discovered that the FAST heading for the Jewish Holocaust is (still) in X11 instead of X47?

Answer: FAST headings can be reported to fast@oclc.org. Because the x47 is a relatively new field we have not necessarily converted all of the x11s that should now be x47 fields. There are not x47 event heading fields in the LC Authority File, so we also have to wait for that if meeting headings haven't been changed to event headings.

When we report possible mis-merged records how long does it take for someone to look at them?

Answer: If you would put something in the subject line to alert us to it being a possible incorrect merge, that would help as we do have several hundred requests that come into our inbox on a daily basis, so we do have to prioritize those requests. So, if you could give them some sort of indication that would get our attention, once we see that it usually only takes a matter of a few hours to recover that merge.

Follow-up: is it better to email than to do an online error report?

We have no preference for us, they all come to the same place, the inbox that QC staff work from. But just based on what was just said about drawing attention to it, an email might be better so you could put something in the subject line.

"National Level Full and Minimal Requirements" is current only through 2010. Is that being updated somewhere other than <http://www.loc.gov/marc/bibliographic/nlr/>? How does OCLC determine requirements where they're not in official MARC?

Answer: That's a question you should ask the Library of Congress. We check that page on a regular basis to see if it's been updated. The Library of Congress has not updated the national requirements for almost a decade now, that would be nice if they would. Since then, we have basically made them up as we go along. We have tried to determine in our QC Policy meetings and in documentation meetings for BFAS what would be the most logical requirements for both full-level records and minimal-level records for everything that's been implemented by OCLC, all the bibliographic elements since 2010. So, we essentially create those requirements which are in the Input Standards sections of each bibliographic BFAS page.

As a Canadian library am I allow to add Canadian subject headings to the records with our library symbols attached to them in Record manager?

Answer: Yes, there is no reason not to.

How are the ramifications of processes for your users and our users being run considered? For example: a process run not too long ago, perhaps because of MARC headings changes, changed the order of subject headings so they are no longer in order of importance as assigned by the cataloger. They now seem to be in order by field numbers. This is unhelpful to our WorldCat local users and probably to other catalogs as well.

Answer: This is a known issue that has been reported. It's happening through some ingest processing where subject headings are being reordered and we have alerted the staff and I believe there is a problem report for that and hopefully being worked on soon.

We are aware of it, it's not an ideal situation. You are more than welcome if it fits into your workflow to reorder those appropriately. We are actively working to try and prevent that from happening, because as you said yes, it is unhelpful to both WorldCat local users and those who are downloading those MARC records.

To go back to the beginning of my question. What safeguards are in place to prevent problems like this subject headings order? We seem to have this kind of problem crop up every now and then and it would be nice to prevent it from taking place at all.

Answer: We 100% agree with you. I don't have a good answer to this other than we are looking at it. The factors involved is whether or not basically is, the record is being brought in through ingest.

What kinds of safeguards is OCLC putting in place is my question, not just the subject heading orders, the process?

Answer: That's part of our development work. We have reports out to our development team, and those have to fit into all the other development processes and priorities that are going on. We've outlined the safeguards that need to go in, talked with and communicated this with our development team. Now it's just a matter of when that can be scheduled to be put in.

We have quite robust validation rules and processes that are continuously being updated and which is one of the reasons why we like you to report any kinds of problems that you run into with the records because if we can find a pattern we can add it to our validation rules, to prevent them from happening in the first place.

I guess I'm confused about the x47 fields. Are a 647 to replace a 6xx field or are both added to the record? In the slide the 647s were fast headings. I thought those were machined generated. Are we supposed to be adding them? If we derive from a record that has fast headings should we keep them or remove them and let the system re-add them?

Answer: As of the time of our implementation of the 647 field, you are allowed to add 647 fields to bibliographic records. Whether there is an LC authority record that corresponds to what would now be a 647 field is a different question and I do not know if there are any yet. Yes, the FAST headings are machine generated although there are institutions which create their own as well. Either manually or through a macro of some kind. You may add 647 FAST headings if they are appropriate, which do not necessarily have to refer to the LC authority file. You aren't obligated to add them, but you may. The system will regenerate FAST headings once a month, so if you are changing LC subject headings in a record of any kind, you can allow the system to regenerate the FAST headings, you don't have to remove the FAST headings or do anything with them and they will be taken care of within a few weeks in most cases.

Hi. I'm creating records for four books in a publisher series. Is there any reason that I can't class them as Pt. 1-Pt. 4, even though each has its own record under a title main entry? And, since this is my first time attending, is this an appropriate question for this forum? Thanks

Answer: Yes, absolutely a perfect question. Also, you don't have to wait for one of these to come along since our next one won't be until January. You can always email questions to askqc@oclc.org, but to answer your question...

You have the choice of cataloging that 4-book set as a set, in which case I would think that the set would have one classification number, and you could do as you wish locally with parts 1 through 4. Its also legitimate to catalog each of those volumes separately if you wish and of course that would allow you to also "classify" them as parts 1 through 4. That kind of issue would be a local determination.

In regard to the question that is being discussed right now, "couldn't this be a whole part instead of series".

Answer: Yes. If there are additional questions on that feel free to email askqc@oclc.org. A lot of times it is best for us to look at the actual records in question to give you a more definitive answer.

When Dewey numbers are discontinued or updated in Web Dewey, how can we view documentation on this? For example, 793.932 is "good" in DDC23 but not WebDewey. Or 641.56362 was added, but not in book. How are catalogers notified when changes like this are made? Thanks.

Answer: At the top of the WebDewey page there are a series of orange buttons, one is labeled Updates which may be the information you are looking for. Or, on the upper right-hand side of the page there is a "contact" button where you could ask this question. There is also the Dewey Blog (<http://ddc.typepad.com>) where you may find additional helpful information.

If there is \$e pn in the 040 could this code be used as a help in validation process? it wouldn't get everything but some?

Answer: This is a good question. Right now, we aren't doing anything with validation from provider-neutral coding but is something we could look into.

This is a problem children's librarians face where the series title is entered with a part title in the 245 or the title is just in the 245 and the series is in the 490/8xx.

Answer: There are differences in [local] practices and whether you do each record separately or one record for the entire thing.

Sorry, no it's not one record for the entire thing. it's how the 245 is entered. As [series title].\$n [number], \$p[part title] or just [title] with 490/8xx {series title} ;\$v[number]. I think this was what the first questioner was asking.

Answer: So we have a lot of variations in cataloging practice here that is not easily reconcilable, and I understand that it's difficult for different communities when they go in and find 3 different records for the same thing.

I understand that MARC changes are announced through reports and bulletins. Is the publication of these reports and bulletins announced anywhere (e.g. a listserv)? How can I be sure to get regular notification of these reports/updates?

Answer: Just last week the Library of Congress issued the new MARC21 Update number 27. If you subscribe to the Library of Congress MARC discussion list you will get all of the MARC update

announcements, all of the technical notice announcements, and so on. You can subscribe to that via the Library of Congress MARC Standards page.

As far as OCLC is concerned, our OCLC MARC changes are announced via the OCLC-CAT discussion list, lists for specific communities such as the Music OCLC Users Group discussion list, the Online AV Catalogers discussion list, etc. And now there is a page that is on the OCLC website which has the WorldCat Validation Release Notes and Known Issues. This page is fairly new and there does not appear to be a way to subscribe to the page.

Is OCLC still merging duplicate bib records for HathiTrust and GoogleBooks? Is it acceptable to add archive.org links to those records since those are accessible?

Answer: Yes.

Why aren't batch loaded records run through the validation process?

Answer: Actually, they are run through a validation process, it's just a different validation process that's not as strict as online validation, because that would possibly prevent a lot of records from being processed through matching and getting added to WorldCat. There are some safeguards in place where really serious validation errors prevent a record from being fully indexed and then they have to be manually corrected in order to be indexed and available in WorldCat. So we have some safeguards in place that keep really corrupt records and records that aren't structurally correct from being added to WorldCat, but some levels of validation errors that we deem to not be as serious do make it through the batch load process, so those records can be processed and then eventually they can be corrected through other means.

I get records that won't validate when I did not make any changes, so I end up deleting fields that are hanging it up but I know nothing about.

Answer: Your options would be to update the record so that it passes validation by deleting those fields, but we hope you aren't losing data and those fields are really illegitimate.

You can also set different levels of validation for when you replace or update a particular record in both Connexion client and browser. We don't know if this is an option in record Manager. Setting the different levels of validation may allow you to do what you want to do without you to messing around with a record you haven't made any changes to.

By all means, if you have any errors you aren't able to correct you can always send them to bibchange@oclc.org

Note: in Connexion client, this is located in Options... under the Tools menu option, under the General tab. There is a selection for Validation Level Options.

I would like \$I issn to be searched as valid with \$a This way I wouldn't need to cheat and use \$y.

Answer: If I remember correctly, indexing of the field 022 subfield \$I simply has not been implemented yet, but it is on the list of things that need to be changed within indexing. As mentioned in today's presentation, the indexing schedule is usually much longer, sometimes long after the MARC update to which it corresponds.

I found a record for a 100 disc set that has 100 020, 024, 028 fields and freezes my session. #885361932. Can that be fixed? We found it with an ISBN search.

Answer: We were able to bring it up, but it locked us up too. This would be something we would need to look into further, but it may be a situation where Record Manager has a better feel for it than Connexion.