January 2020: Best practices for editing WorldCat bibliographic records

Topic-specific questions

January 7, 2020

Can you please review which type of hybrid records should be reported, rather than edited?

For language of cataloging hybrid records, both print and electronic records, you’ll want to report those which have close to, or an equal amount of, descriptive elements in both languages. Or records that have pretty close to the same number of holdings based on location. Any time you aren't sure what a record represents, feel free to report those to us at bibchange@oclc.org.

How can we tell the language of cataloging of libraries with holdings attached? Do we need to check each library?

It depends on which interface you are using. In Connexion client, you can select the View menu option, then Holdings..., then All. This will display the holdings by country, state, province. Looking at how the holdings are grouped; you can get a general idea which language of cataloging likely has more holdings based on where those libraries are located. In Record Manager, holdings can be displayed by clicking the link stating how many libraries hold this item from the list view, or by clicking the other libraries link containing the number of holdings while viewing the record. There are tabs for All WorldCat Holdings, Holdings in My Region, and Holdings in My State.

Approximately how long does it take for OCLC to fix the language code mix-ups once reported?

Bibliographic change requests reported to us at bibchange@oclc.org are placed into staff workflow. requests are processed first in, first out, so depending on the number of requests received, Generally, requests are processed within a week.

Is there a way to submit errors to OCLC from within the bib record itself?

Yes, there is. In Connexion client, if you have the bibliographic record open, select Report Error ... under the Action menu item. This will open a dialog box that will provide us your OCLC Symbol, name, email address, and a free-text box for you to describe what you are reporting. A copy of the record as it is displayed on your screen will accompany the error report. In Record Manager, with the bibliographic record open, you can select Send Record Change Request under the Record drop-down menu item. In the dialog box, you can select the type of requested change, provide a description, your name, and email address. In Record Manager, a copy of the record is not attached to the
request we receive. OCLC's Bibliographic Formats and Standards, Chapter 5.5, Requesting Changes to Records documents the different ways you can report errors to us.

If I find a record with no language of cataloging indicated (i.e. no 040 $b) but with all the notes & 3XX fields written in French, should I add the French lang of cat code or report the record to OCLC?

Yes, you can add the field 040 subfield $b. There might be a few of these in WorldCat but should be fairly rare to come across. We have been making progress making sure records coming into WorldCat have a code in the 040 subfield $b.

For CIP records, I missed the part of 040 that cannot be edited. What subfield can't be edited again? The $b?

Yes, the 040 subfields $b and $e should not be edited in CIP records.

Bib Format & Standards for 588 field says: Note: Note: CONSER participants should not use 1st indicator values 0 or 1 in CONSER-authenticated records for integrating resources and serials at this time. The Library of Congress and OCLC will announce the implementation of these values for use in the CONSER database at a future date. Is this still the case?

As far as we are aware this is still the case. We do not have a date yet when we can change this practice. When that does happen, we will be looking at converting existing 588 fields to conform to the way the rest of them mostly look in WorldCat now. Currently, we are waiting for the Library of Congress to make an announcement to when they are making their system changes so that we can update documentation and have CONSER participants code field 588 with the indicators.

When making a new record, do you encourage to add e-ISBNs to a print version record in 020 $z, and vice versa?

Yes, that would be fine to do.

Usually when I edit a record, everything goes well, but sometimes I get a message saying I don't have permission. What determines that?

The level of cataloging for your library or institution can determine what you can and cannot edit in a bibliographic record. In Record Manager, this is called Record Manager account roles. OCLC's Bibliographic Formats and Standards, Chapter 5.2, Member Capabilities documents the types of edits which can be made based on the different authorization levels. Next month's Virtual AskQC Office Hours session will cover enriching bibliographic records and will also cover this topic.

What are your thoughts on removing 650 like "Mexican fiction--21st century" for individual novels by an individual author, which Subject Headings Manual explicitly tells you not to do, but a lot of libraries seem to like adding to records for their local users?

Subject headings that violate the instructions in the LC Subject Headings Manual are okay to remove.

You said it's allowed to edit the date in the call number in the 050. Is that only allowed for ELvl 8 records? With the piece in hand?

With the piece in hand, you may edit the date in the call number. This is also true for other records as well, not just encoding level 8 records; however, you may or may not be able to edit field 050 of records created by a national
library depending on you level of authorization. Please refer to OCLC’s Bibliographic Formats and Standards, Chapter 5.2, Member Capabilities for more information on this topic.

**When deriving a record in WMS-RecMgr, the 040$e 'pn' (for provider neutral) is lost. Is there a reason for this or is this a bug?**

The 040 field does not transfer when deriving a bib record. If the 040 $e rda appears in a newly derived record it’s due to a user preference for creating new records with Default Record Format = RDA.

At Library and Archives Canada, we assign both English and French subjects regardless of language of cataloguing. We would like to make sure that subjects in both languages are not deleted from records based on the language of cataloguing in 040 $b.

Subject access points are not considered when determining the language of cataloging. Records can have subject access points from different schemas and in different languages. These should not be changed or deleted from bibliographic records. For more information, please see Bibliographic Formats and Standards, Chapter 2.6, Language of Cataloging.

**What is the latest news regarding the new guidelines for minimally punctuated records? Are macros available yet for adding or removing punctuation in OCLC records?**

Yes, there are macros available now which will work in the Connexion client. It is part of the OCLC Macrobook 2019 which can be downloaded from the Cataloging software downloads page along with the OCLC Macrobook 2019 instructions. Laura and Robert have been on the task group to work on minimal punctuation guidelines training materials which are just being finalized and those should be made available on the PCC website, probably sometime this month. We will be having a Virtual AskQC Office Hours presentation on this topic in March.

**Do records with encoding level 8 and I get merged at some point?**

They can, but what often happens is encoding level 8 records get upgraded by either the original library or by a PCC participant. If you find cases like this you can report them to us as duplicates.

**Would you expect cataloguers in the UK to use 'color' in, say, field 300, or can we get away with use of UK spelling 'colour', particularly when creating records from scratch?**

It is a valid spelling, so if you are creating the bibliographic record, you are certainly welcome to spell it colour.
January 16, 2020

This is pretty hard to fix records in an Asian language you don't know when the only English element is p. and cm. in the 300.

If you don't feel comfortable or have the language expertise to edit a record, it is best to leave the record as is.

In the French hybrid record example wouldn't the 300 field also be in French?

Yes. The word 'pages' is the same in both English and French.

For books in French, would the 245 be in French, not English?

In the presentation example, the text of the resource is Italian, and the code in the fixed field in Lang is 'ita'. The code in the 040 subfield $b is 'fre' as the bibliographic record was created by an institution that catalogs in French.

What documentation indicates that we don't add extent for online resources? There is an example in BFAS that includes page numbers. I know you're not supposed to include the size but the last quiz indicated that adding the page numbers was incorrect.

Extent is appropriate where applicable in bibliographic records for electronic resources.

On the last spot the error quiz, the answer options were: A) Field 300 includes size  B) Field 337 has incorrect term  C) Field 856 has incorrect second indicator  D) All of the above.

How can you tell when you can change a record coded as print into electronic resource and when are you not allowed to do that?

It is hard to supply a definitive answer that could apply to all cases. You would want to evaluate these records individually on a case-by-case basis. If you have questions about a specific record you are always welcome to ask us at askqc@oclc.org or bibchange@oclc.org and we will be glad to work with you.

For English-language cataloging agencies holding foreign-language materials for patrons who speak that language, how much can/should you change a record before it becomes "hybrid"? You may have, for example, subject headings in both languages.

Records should only contain descriptive cataloging data in one language. If you wish to add cataloger-supplied notes for non-English speaking patrons, you will want to use fields for local use for that information. For 505 table of contents notes and 500 quoted notes, these are taken from the resource and should be in the same language as the resource.

The language of the subject access points is not a factor when determining the language of cataloging. Use the 6xx fields to provide subject access entries and terms. Most 6xx fields contain subject-added entries or access terms based on the authority files or lists identified in the 2nd indicator or in subfield ǂ2 and terms can be in languages other than the language of cataloging.

For hybrid non-English language of cataloging records, could you explain further about considering the number of holdings by English language of cataloging libraries? Is there a tipping point where you would change the language of cataloging?

Holdings are only one factor to consider when evaluating a hybrid record and there is no tipping point where based
on this information alone you would change the language of cataloging. You will also want to determine the intent of
the cataloging agency and the number of descriptive cataloging elements in different languages. As a general rule, a
record's language of cataloging should reflect the predominant language used in the descriptive cataloging elements.
For more information about language of cataloging and hybrid records, please see Bibliographic Formats and
Standards, 2.6 Language of Cataloging.

If the book I have in my hand has no date but is probably published after 1945 and the OCLC record has 1991?
what date should be used for the call number, would we add a z to 1950z.

Assigning call numbers would be a local practice and would depend on how your institution assigns call numbers or
revises them for local use.

You're not supposed to put the 505 and 520 in the language of the resource if you're an English language
agency?

The 505 contents note is taken from the resource and should be in the language of the resource. A field 520
summary note is cataloger supplied and would be in the language of the cataloging agency unless a summary is
being taken from the resource itself, in which case it would be in a quoted note as it appears in the item.

When cataloging an academic thesis will we still use the 500 tag to “include appendix (leaves 54-60)”?

Yes, an includes appendix note or index only note goes in field 500, they would go in field 504 when combined with a
bibliography note.

For videos that have no language (no audio videos) would it be okay to remove the "eng" in 041?

If Lang (field 008/35-37) is coded zxx (No Linguistic Content), then you most likely would not have a code in field 041
subfield $a. If there are credits or accompanying material, then there may be language codes used in other subfields
in field 041.

Is it okay to add annotations in both languages if an English cataloging agency catalogs books in Spanish,
French, etc.

This would be a situation where the MARC standard doesn't necessarily support the use of multiple languages for
describing a resource. So, if you are an English language institution you should not add non-English language
descriptions unless you are able to add a quoted note from the item.

Why can't full members edit encoding level? I often find very complete records with low encoding level. I
assume this is because people have enhanced them, but they can't edit encoding level. What are we supposed
to do if we enhance a record like that?

You cannot change encoding level 8 or encoding level 1. Next month's Virtual AskQC Office Hours session will cover
enriching bibliographic records and will also cover this topic.

But you can change ELvl M to I?

Yes, you can change encoding level M to I.
What would be the criteria for either merging or not merging records?

There are a lot of things taken into consideration regarding merging records. If you look at BFAS, Chapter 4, When to Input a New Record you will see many of those criteria are essentially the same criteria used for merging or not.

Hello--I recently noticed that the WorldCat.org website sometimes has Table of Contents from machine-generated Table of Contents, if there is not a 505 in the print record. Do you know anything about this procedure and where the TOC are coming from?

They are being imported from third-party vendors along with abstracts and summary notes when lacking in bibliographic records. If a cataloger adds a table of contents note into the bibliographic record, that will replace the imported note being displayed in the WorldCat.org interface.

I just reported a record the other day that looked like it should be merged. Is it better to report an error in Connexion or email? Or doesn't it matter?

OCLC's Bibliographic Formats and Standards, Chapter 5.5, Requesting Changes to Records documents the different ways you can report errors to us. Whichever method works better for you and your workflow is the best method to use. All methods come into the Metadata Quality staff workflow for processing.

Is there any kind of notification process for when records are marked for deletion or when they are about to be merged?

No, there is not.

We recently notice tag 227 used to record another OCLC record number. Does this indicate a duplicate record?

No, it does not, field 227 is not a valid field in MARC. Thank you for reporting this to us so we could locate them and delete the fields. Records which have been merged have their OCLC Record Number (OCN) retained in the 019 field.

February 2020: Best practices for enriching WorldCat bibliographic records

Topic-specific questions

February 4, 2020

Is adding a series access field considered a change to the earlier cataloger's local decision?

No, the cataloger may have decided not to add it or the vendor or publisher may not have had that information available at the time the record was created.

Could you please comment the best course of action for institution-specific URLs in Kanopy WorldCat records? Should they be deleted or something else?

Institution-specific URLs should not be added to a WorldCat record. They should be entered in a local holdings
record (LHR). We encourage removing these types URLs. When this problem is reported to Metadata Quality via bibchange, sometimes we are able to find a generic URL to replace it with, but not always.

Several of the enhance examples showed adding an additional 264 field for copyright date even though it was the same as the publication date. Is that something OCLC recommends?

That is totally up to you, it is optional. LC practice is such that they do not do that if it is the same as the publication date. But you are free to add it. We have both, plenty of records that do and plenty of records that do not. They are both correct.

Looks like it’s now okay to import diacritics to WorldCat records without having to overlay with Connexion diacritics, is this correct?

It sounds like this question has to deal with the case of the pre-composed letter diacritic versus decomposed where the letter and diacritic are entered separately. If picking up stuff from a web source where you really don’t know the nature of that character, it is probably pre-composed. Yes, that is no longer a problem as it was before the day of implementing all of Unicode. You can put them either into the Connexion client, Connexion browser, and Record manager.

What is generally the turnaround time for issues submitted to bibchange?

Generally, we try to turnaround things in one week, but please note there are some requests that may take additional research, so the one week turnaround may not be possible.

If a record is in the database cataloged under AACR2, should you enrich the record to RDA or leave the WorldCat record as AACR2?

That is your choice. Both are acceptable. There is no requirement to do that.

Could you please clarify the difference between Replace and Update?

Update is when you are setting your holdings. Replace is the command that replaces the WorldCat record. The command Replace and update will do both.

Could you discuss making changes to AAPs when differentiating personal names previously on undifferentiated NARs?

If you want to differentiate names that were previously undifferentiated you are welcome to do so. To do this you could use as differentiation birth and or death dates, or something to describe the individual like a title or occupation. Unless you are creating a name authority record and work under PCC rules, you are not required to differentiate names. In addition, if you don’t participate in NACO and you are aware of information that you could share with Metadata Quality to differentiate a name heading, you can send the information to authfile@oclc.org and share the information you have. We can probably create an authority record to represent that differentiated name.

Can anything be done on the automation end to stop people from converting a monograph record into a multi-volume record instead of creating a new multi-volume record?

That would be a great thing. We haven’t come up with a way to do that. If you have ideas for us we would be happy to listen. You are correct it is not good practice to change a single volume into a multi-volume record. Also, you can
always send these issues to bibchange@oclc.org. We would love to get any examples you have with seeing single volumes being changed into a multi-volume. We can look at the history of the record and change it back.

**What is the policy on 336/337/338? Is that a valid reason to edit the record - and would that be an enhancement or enrichment?**

To reemphasize what we went over in this presentation, enhancement is when you are changing the encoding level on the record from one to another. If you are adding 33x fields, that would be considered enrichment because you are adding fields to the record. If you do other changes to the record to make it fuller, then you would want to change the encoding level, which would then be an enhancement. In Metadata Quality when we run macros and add those 33x fields to the records, that’s completely valid. It’s okay that we have a hybridized database with elements such as those in AACR2 records. So yes, you could definitely add them if they are not present.

**Occasionally when I try to replace a record after enhancing the contents note, the replace fails because of the presence of the last field, which says DAT; the error message says that the field is not valid (I forget its tag) What is that about?**

To be honest, that is not ringing a bell to anyone in the room, so if you would please send an example of the record that introduced that error to us we could do further investigation, but that is not sounding familiar to anyone.

**What is the best practice for recording $c in 520 field? When do we use $c and when don't we?**

We are struggling to remember what the $c is. I think that is when you say where the source of the note is from. You would only use that when you are quoting that note or taking the note from another source rather than composing the note yourself. What is the best policy on that? It really is at your discretion whether you use that or not.

**I sometimes see data in 490/830 series fields of WorldCat records that reflect local data such as "John Smith Alumni Collection". Should I only delete those locally?**

I would say that maybe you should send those to bibchange@oclc.org so that we can do further investigation. If you have the item in hand and can see that it is not really a valid series statement, then it could be deleted from the WorldCat record.

**What is the status of local subjects (6XX _4 or 6XX _7 $2 local) when adding subjects from standard vocabularies? Should the local subjects be preserved when they're identical to a standard subject heading on the record?**

There is no reason you need to preserve a local subject heading that is an exact duplicate of a standard subject heading that is already in the record. Feel free to delete those.

**When we create new NARs, what are your thoughts on how much/what kind of cleanup we do ourselves versus reporting BFM to OCLC and letting you all deal with it? Should BFM be reported to bibchange, AuthFile, or a different address?**

If you have records in your local catalog that need to be updated then go ahead and do it because it is in your local system and you want them to be updated. If there are other records in WorldCat and you don’t have these resources, go ahead and send them to Bibchange because we are dealing with bibliographic records. Also eventually Controlling should update these headings if they were previously controlled.

**This is a follow-up to the question about changing single volume records into set records. Sometimes a**
publisher decides to turn what was originally a single volume work into a set after the fact, i.e., the first piece gives no indication it is part of a set, but then another one comes out calling itself volume 2 and referring to the previous one as volume 1. In that case shouldn't we change a record for a single volume into a set record, or should we create a new record for the thing as a set?

When the initial record is definitely cataloged as a single volume and doesn’t indicate that there is an open date or open volume, create a new record for the multi-volume set. OCLC policy is that it is appropriate to have a record for a single volume in addition to a record for the entire set. So, there could be a record for volume 1, volume 2, and a record for the multi-volume set.

Does OCLC pass BFM requests on to LC?

We report to LC anything that is in their catalog that needs to be corrected, but not normally bib file maintenance.

Going back to the BFM question: I've never reported BFM requests when an AAP in a 240 changes. Should we be doing this if we want them to get changed without our changing them manually?

Yes, since 240 fields are not subject to controlling, it would be great if you reported them. We would not know otherwise.

What's the policy of recording the price information of an item in 020 field?

Please don’t. There is a place to do that in the MARC format but we routinely delete them because they are not generally valid over time. They don’t help anybody to know the true price since prices change and data becomes out of date. Plus, there are currency issues. BFAS says to generally not enter terms of availability except for scores.

Is OCLC considering changing the types of encoding levels?

Yes. This is something we have been considering for quite awhile. Our long-term goal is to eliminate the OCLC-specific encoding levels which are the ones that are alphabetic characters and instead adopting the standard MARC encoding levels, these are the numeric codes. This is something we have been planning for quite awhile and we are still planning. We don’t have a time frame yet. These encoding levels are embedded in so many systems within OCLC that it is taking a long to consider. We are taking time to consider every system and service that may be affected. When we are ready to implement that we will give lots of notice and information on what is going on.

What kind of information OCLC needs to be included in the BFM request?

It is up to you if you want to give us lots of information, but not necessary. We all tend to look at the authority file record to make sure we are changing the titles that should associated with that entity. We could always do follow-up with you if we need more information. It is not necessary to supply all the record numbers because we search regardless, to be sure we are catching all headings that need to be changed.

Are encoding level values factors in doing bib file merges? Do changes to encoding level values prompt other kinds of automated attention?

Yes, when doing manual merges our automated DDR merging program does use encoding level as one of the factors. More important is the completeness of the record. As explained, Encoding level M could be very brief or very full and so a lot of factors get taken into consideration. Does that prompt other kinds of automated attention? Yes. When you are enhancing a record and replace the record, that makes it a candidate for DDR.
February 13, 2020

I was under the impression we couldn’t edit 245 fields or other fields already in the record. Is this a difference between enhancement and enrichment?

Except for PCC records, you are welcome to edit the 245 field in records. Hopefully, if you are editing the 245, it is to add something that is missing. For example, if it cut off after the $b and you wanted to add $c, or if you are correcting a typo. Those would be the type of edits that would be appropriate to make to a 245 field. The difference between enhancement and enrichment is, when we are enhancing we are changing the encoding level to a higher level, such as 3 to a K, or a K to an I. Enrichment is adding additional fields to the record. Editing is actually something that was covered in the January AskQC office hours.

Is enhancing from M to K permitted? I read BFAS 5.2 as M only to full.

It is permitted even though it is not shown on the table. There could be times, as the examples that I showcased in the presentation, that you might not have all the initial information at the time and are changing a few data points that you have. You are only going to change from a M to a K rather to an I because you haven’t fully described the resource at that point. So yes, you can change encoding level from a M to a K.

I see that editing typos comes after replacing the record, does that mean you cannot edit typos in PCC records?

You are welcome to do what we call local edits when fixing typos in PCC records prior to adding your holdings to the record. Unless it is one of the fields that you are allowed to edit as shown in the chart in BFAS, you probably won’t be able to correct the typos in the WorldCat record. That is when it would be very good if you would report those errors to bibchange and we will make the changes for you.

If one’s library is PCC, but only one cataloger has been BIBCO trained, is it still OK for all to change a PCC record?

If you are enriching the PCC record by, e.g., adding a contents note that was not there, anybody with a full level authorization is permitted to do that. If you are doing other changes that aren’t permitted for everybody, it should be done by the person who has the BIBCO or National level enhance authorization.

May I change wrong formatted diacritics in non-English records?

If you have the language expertise that would be fine. If you do not it would be preferable to send a request to bibchange.

Can you talk about OCLC’s project to do non-Latin enhancement of records?

So far, the only language we are tackling is Russian. These are records that are cataloged in English where the language of the item is Russian but lack Cyrillic characters. This is a project that is cooperative between OCLC Research area, Metadata Quality area and UCLA, who was doing a similar project, so we joined forces. We have started the work. We are supplying Cyrillic fields for those records. We are enhancing the records with those non-Latin fields. If this work goes well, we may expand this to other languages or scripts. We need to get this work done and evaluate what was done before we make any choices to move on.
How long will the Russian project take?

We just started the replacements to the records this month. We think we may be done as early as next month.

How can we tell if the Cyrillic was added by this automated project?

We are adding a specific note field to say that we have done this. I don’t remember exactly what that note is.

It may be a 588 field, an administrative kind of note, that catalogers would be interested in, that will say that it was part of this project. That way if a cataloger was reviewing it in the future with the item in hand, they could review the Cyrillic that was supplied and decide if it looked okay or correct it. Then at that point, get rid of the administrative note. It was meant to be an alert to other catalogers that there may be a problem with the Cyrillic that was supplied. It was all dependent on what the transliteration looked like.

What will be in 040

I think it will be OCLCQ but I am not positive.

When I request an edit to a PCC record, what is the typical turnaround time before I'd see that change?

Usually we can process requests within one week.

I have seen adult level materials with juvenile subject headings (6XX _1). Should we delete those headings in the WorldCat record?

That is appropriate, if the juvenile subject headings are not correct for the resource.

Why does it take so long to merge records? I've seen records that are the same for the same manifestation, and they haven't been merged!

We have different ways of merging records, one is DDR, our duplicate detections software. Jay Weitz has given presentations on that. We err on the side caution with that because we don't want to inappropriately merge records. Manual merges are very labor intensive. If they are not reported to us we don't know they are there. We do come across many in our daily workflow. We also have a large backlog. We are doing different initiatives with the Member Merge Project to help with this big duplicate problem. Thankfully, many libraries are helping with the issue we are having with duplicates. Please keep reporting duplicates and we will get to them as soon as we can.

We have two distinctly different backlogs in terms of reports, one is the Bibchange for changes to records, the other is duplicates. We try very hard to keep the backlog to a minimum for changes to records, that is a turnaround time of a week or less. The duplicates take a lot longer. Sometimes we can do them fast, but it depends on the other work that is going on, or what other things have been reported. We do have a large backlog of duplicates. Some formats are caught up but books we have at least a year’s worth.

I understand the natural fear of merging the wrong records. Is there is documentation on what fields the merges are based on?

We have internal documentation that we have shared with our Member Merge library partners. It talks specifically about what fields to compare. It is very similar to the documentation that is in BFAS Chapter 4, When to input a new record. So, if you look at that, you will see what we pretty much use as the same criteria.
A previous webinar said an OCLC member could apply to merge records, is that still possible? Would love to do that for all those level M's.

We would love for you to do that as well! You have to be a member of PCC in order to participate in the Member Merge Program. We will be starting a new round of institutions for training in August or this fall sometime. If you are interested send a message to AskQC@oclc.org to express your interest.

Can we change the 050 field call number when it doesn't represent the work well, maybe only partially? I see records with two 050 fields.

On non-PCC records, anyone can add a 050 second indicator 4 which indicates the call number is locally assigned. It is not a good idea to change 050 with second indicator 0 on the LC contributed records, but any with second indicator 4 are fair game. If you do see an error on a PCC record and it is 050 with second indicator 0 from LC, you can report those to bibchange@oclc.org. We can look into it, make the change, and then follow up with LC.

Is there any financial credit for either enhancement, or enrichment?

No. There used to be but OCLC eliminated charges and credits years ago.

In the example of the audiobook being enhanced from K to I that had the 347 audio file and CD audio in the same field. I thought terms in the 34X field were supposed to be recorded in separate fields, especially if they come from separate lists, i.e. audio file is from the file type list and CD audio is encoding format.

Yes, that is correct they should be in separate 33x fields if they are from separate vocabularies.

To whom I send an email to check for a merge suggested last year and it hasn't happened?

You can resend that to bibchange and we will look into it.

If we think a record has been merged incorrectly can you tell, and how long after a merge before an 'unmerge' can happen?

It depends how far back it was merged. We may be able to see what a record looked like before it was merged as far back as 2012, anything before that we cannot. If we can’t see its history in our system we may be lucky enough to be able see it another library's catalog. They may have a copy of the original record. Any record that you think was merged incorrectly please send to bibchange. If we notice a pattern of things not being merged correctly, we can pass it on to the folks that work with the DDR software and can fix it. We merge a lot, so please do report problems.

Any development regarding the compound characters with their decompound form in OCLC like the dot below with certain languages?

This is an issue that has come up for us over and over. Particularly the way that these characters in their pre-composed form can be a problem for some catalogers only because some of our tools are a little bit older and are not completely Unicode compliant. Several years ago, we implemented all of Unicode and suddenly you had the possibility of representing a letter with a diacritic in two different ways. That’s the issue that plagues us at this point. We started a discussion of how we could deal with these in the longer term. We are going to move forward in taking that issue to our developers and see what kind of solution they can possibly suggest.
Why doesn't BFAS 2.4 link to the RDA and AACR2 full-level standards?

RDA is subscription based. So, unless you are a subscriber a link would not work. The online version of AACR2 is part of LC's Catalogers Desktop which is also subscription based. We don't because not everyone has access.

When we add additional LC subject headings to a record, how long does it take to have the FAST headings generated? If a cataloger adds the FAST headings, will the same FAST headings be generated by the system?

It is normally within a few weeks. I think there is a monthly cycle generating records that are new that do not already have FAST headings on them. It depends on when a record comes in, whether it was entered manually or by Batchload, so possibly 4 to 6 weeks.

When we see new materials cataloged in AACR2 is it always correct to replace it with RDA standards?

You are welcome to do that, but there is no requirement.

If we derive a new record from one with fast headings, should we remove the fast headings in the new record and let them be generated again?

If the headings are applicable, e.g. you clone the record for the second edition to add the third edition for the very same title, and you are going to leave the same Library of Congress subject headings, then the same FAST headings are going to be generated. You can leave the fast headings there in that case. If you are going to change the Library of Congress subjects, then you should delete the FAST headings and let them be regenerated.

Is it wrong to derive from existing records instead of starting with a workform? I tend to find something close and derive to save typing.

It is an easy way to do that you but you need to be careful that you did not miss something. It depends on the cataloger; some like to start with a fresh copy also.

March 2020: Punctuation updates and policies

Topic-specific questions
March 2, 2020

Can you remind me again what these record examples would be coded as?

If you are creating a brand new record and you are including all of the punctuation, there is no difference in how you would code that over what you had been doing in the past. The difference would be in coding Desc in the Fixed Field if you decided to do minimal punctuation. So, for RDA, you would code Desc: c rather than Desc: i which indicates that the punctuation has been omitted. Also, if you were coding a record as AACR2, you would code Desc: c rather than Desc: a and then add an indication that AACR2 rules were used by putting in 040 $e aacr/2.

Can you please give a brief review as to why removing the punctuation is a good idea? What is the benefit to the users?

The aesthetics of it. It's easier to probably supply punctuation in some display schema than it is to suppress all of the punctuation in a display that we sometimes do not need. So, to make the data easier to manage, it would be better if punctuation wasn't there. How you view this is very much governed by your local system and its capabilities. In these discussions over the past several years, people are either okay with this idea or they really don't like it at all. We can appreciate both viewpoints based on what system you have to work with. We also understand that the BIBFRAME to MARC conversion that is now being developed by the Library of Congress will most likely omit the punctuation. So, when it's converting from BIBFRAME back to MARC, the MARC record that results won't have punctuation. Like it's described here, it will omit the final punctuation and the medial punctuation. That is another reason that, going forward into the future, we'll probably be seeing less and less records with punctuation.

Should punctuation be relocated (e.g. 245 $b) in all cases?

In the case of 245 $b, it's just the equal sign ( = ) indicating parallel language information, and the semicolon ( ; ) in the case of multiple titles without a collective title. If it's just the ordinary $b where you subtitle or other title information, that colon would be omitted altogether rather than relocated.

Is there a space before and after subfields? For example, space $b space?

In the case of OCLC displays, there is spacing around the subfield codes. That is not actually the case in what would be output as the MARC record.

So, if we create a new record, should we use this new policy or is it optional?

It's optional, whatever works best in your system and in your situation. If you prefer to continue using punctuation, that is fine. If you decide to omit punctuation in step with these guidelines, that is certainly okay as well.

In an RDA record with Fixed Field Desc: i, sometimes one or two pieces of punctuation are missing. Other than final punctuation, should these be supplied for consistency and the WorldCat record replaced?

It is certainly okay to do that. The punctuation in the record should be consistent with the coding in Desc. If it's the case where most of the fields have punctuation and the description is coded as Desc: i, then the intent would be to go ahead and include punctuation in fields. That does not necessarily involve terminal periods. If you have a record that is coded Desc: i and you notice that the colon is missing before 260 $b, go ahead and put that colon in and replace the record.
When leaving off final punctuation only, will the code remain Desc: i or Desc: a?

Final punctuation is not the determining factor in how you code Desc because it's optional for both records that have full punctuation and records that have minimal punctuation. If you are including most of the punctuation as you have routinely done and you're just omitting the final periods, it would still be coded as Desc: i rather than Desc: c, in the case of an RDA record.

For the following example, which punctuation can be omitted?

245 00  Ji qian jia zhu Du gong shi ji : $b er shi juan ; Du gong bu wen ji : er juan.

In terms of what is provided in the example, the colon before $b could be omitted, but the other punctuation that occurs within the $b would be retained.

My reference librarians are concerned that the colon separating a title from the subtitle would create inaccurate citations using citation software. Apparently some citation software relies on the colon to create citations. Does OCLC include the proper punctuation in the citation software?

OCLC does not put out our own citation software. There are a number of commercial packages or other home developed packages of citation software that are out there. You are welcome to use those with OCLC records, but we can't predict what will happen with those.

I've seen a few records recently with commas in the place of the ISBD punctuation (e.g. in the 245). Should these be removed or replaced?

Given what our policy is at this point, you would want to replace those. In other words, fix the punctuation rather than go the route of removing it. If you have record that really is a toss-up because there was some coding in Desc that really didn't match the coding in the various fields, some had it and some didn't, but it's half and half. At that point, it probably doesn't matter which way you change the record. The goal would be to make it consistent. Incorrect punctuation should always be corrected if you notice it and want to do so. This might be a good example of why it might be better to omit punctuation than to fiddle around with fixing it.

Are there any plans to eventually remove punctuation from existing OCLC records created before the approval of the new minimal punctuation policy?

This is something that we would like to do, but it is a long-term strategy. We haven't determined whether we will or will not do it, and if so when we would do it. It is something we have talked about internally and would like to do, but it does not mean that we are going to do it. Our reasons would be that we think the data, going into the future, will be more consistent. If we did, we would provide advance notice of our doing so, so that you could deal with records that come to you as an output from Collection Manager or just the fact that you would find more records in the database that lacked punctuation.

If the punctuation can be removed automatically, why would individual catalogers take the time to punctuation from existing records?

That would depend on local policies, workflows, and how much time people want to spend on records. This would be a prime place to use the OCLC PunctuationAdd or PunctuationDelete macros.
WorldCat Discovery removes all punctuation. Isn't that an advantage of Discovery?

Certainly, those of us who are in favor of removing punctuation think it's an advantage. When we have talked to our Discovery colleagues here at OCLC, they think the idea of removing punctuation from all bibliographic records is great because they won't have to programmatically remove it for displays in Discovery.

Is the Fixed Field Desc: i or Desc: a used by any public catalog you know of?

It's really probably more of an indication for other subsequent users of that record in terms of cataloging. I don't know of any catalog that takes that code and then, as a result, does something special in the display. The code Desc: c to indicate that it's an ISBD record with punctuation omitted was added to the MARC format somewhere around 2008-2009. It was a proposal made by the libraries in Germany. We load a lot of those records, and they have no punctuation in them. The value of that code is to indicate that is how the record is supposed to be, so that if punctuation snuck back in to such a record, it could possibly be removed. In the case of Desc: i, if a field was added to the record even through field transfer that we do here at OCLC, we could say that field is supposed to have punctuation and possibly supply it.

If we encounter a record with inconsistent punctuation, should we make edits and replace the WorldCat record to remove the punctuation?

Yes, there is no reason that you couldn’t go ahead and do that if you want to. Like anything else in a record, it's up to you whether it's worth the time and effort to fix a problem. A punctuation problem is relatively minor in terms of everything that could possibly be incorrect in a bibliographic record.

Will Discovery or any discovery system you know of be able to determine whether ISBD moved to the other side of $b such as = is ISBD vs. integral to the other title info?

As there are more records that have minimal punctuation, systems will adapt. If you are thinking along the lines of if a system is automatically going to supply a colon before $b no matter what, probably few systems know to do that at this point. So if that is programmed into some system display in the future, it would need to take into account 'supply the colon before $b except if there is an equal sign or semi-colon as the character in $b'. It does get you out of the problem of having a display of the title proper in $a with an equal sign hanging on the end if $b is suppressed from a display.

How does duplicate detection affect this? Can records be merged with different standards for punctuation?

That certainly is the case now in that we deal with different styles of punctuation in duplicate detection, so there is no problem in our being able to compare fields and match records when they're created under different rules. If there is a publication from the 1950s where there is no ISBD punctuation but we have a duplicate record that does have ISBD punctuation, we are able to compare those fields and then select the record, not necessarily on which rules were used but more in the case of completeness of the data and the number of holdings that are found on that record.

How will this affect clustering with newer and older records?

As is the case with duplicate detection, it really should have no impact because we normalize out any kinds of comparisons we do with data, any punctuation that's there. So, we are typically looking at the wording in any subfield in order to tell if any records look like they represent the same thing in terms of duplicate detection. Or that they are
different versions of the same work in terms of other clustering.

Some staff think the colon between the title and subtitle makes it easier to read the title display in Discovery. Without the colon it's hard to know where the main title and subtitle begin and end. Is it possible for OCLC to automatically add a colon to the display?

That seems to be the most common question in terms of removing punctuation, but the idea behind is that display punctuation shouldn't reside with the data. Display punctuation is something that systems should supply. So, in answer to the question about is it possible, certainly it is possible. Schedules for implementing these kinds of things are always the issue.

What impact on patrons can you think of for records with or without punctuation? Citations generators is all I can think of.

I'm not sure I can think of anything else. I look at is from the perspective of clean data that can be easily manipulated for various kinds of displays without the need for suppressing that kind of data. It's the case that, because MARC was developed as long ago as it was with punctuation intact, and that we've never made the effort to change that, that we still have punctuation within our bibliographic data.

March 12, 2020
Topic-specific questions

Did you explain why punctuation is being removed? It seems like the 300 field would be confusing without the + sign preceding accompanying material.

The plus (+) preceding 300 $e is not really any different than a colon or a semi-colon that precedes a subfield in other cases. You could always expect that preceding 300 $e you would have a plus sign to indicate that accompanying material is what follows. Which is exactly what $e means, so it's a piece of punctuation that could be omitted.

Another case of transposed punctuation: Dr. Strangelove $b, or, How I stopped worrying .... Correct?

Alternative titles go in $a. So, as a result, the commas that you put around "or" in that case would all be in $a and is punctuation that would be retained. You would also create a 246 with the other title information.

The record examples are 'old' records with punctuation removed for the slides. Can you provide OCLC numbers for records that were created using this standard?

Sure, we can add a couple of current examples to the slides. If you really want to look at some records quickly, the Germans, with German language of cataloging, have been using records with no punctuation for quite a while.

The most current says OCLC Macrobook 2019. Is that the version that includes the punctuation macros?

That is the most current one. It was modified at the end of last year in preparation for the implementation by PCC. It was released in December, which is why it is labeled 2019. If we put out another one to revise those, which we probably will do sometime this year, we will widely publicize that and it will be labeled 2020. You can get the macros from Cataloging Software downloads.

Our current ILS supplies the commas in the main or added entries if the person has two relator terms so these now show 2 commas (i.e. director,, producer). How long do you think it will take ILS’s to catch up and supply the punctuation to make the display more patron friendly? Without the display of the delimiter and codes it’s difficult to read.

It sounds like your ILS supplier is already providing commas and that is clashing with records that have commas embedded in them. If those commas weren’t there, you would have a single comma rather than two. Of course, how long it will take for local systems to catch up is up to whoever sells that system. If they heard from their users they would be more inclined to make changes rather than if they heard nothing from their users.

What is your advice for removing punctuation from enhanced 505s?

Field 505 is kind of a problem. In the macro that we put together, we skipped removing punctuation from field 505. In part because in an enhanced 505 field, you could potentially have $g which can represent a lot of different things. If it’s a sound recording record, $g may contain the playing time, the duration to a piece of music. In the context of a book record, the $g might indicate the volume number. So, it's unclear what should happen if you have a 505 field that just contained subfield codes, what punctuation would be needed for display. This may be a case where the MARC format ought to be modified to make a clearer distinction between how some of these things are used when it comes to field 505.
Topic-specific comments

This looks like Anglo-American Cataloging Rules from the 1960s and 70s.

If you are thinking in terms of it looks a lot like records before ISBD punctuation came along, I would say that's true. But those rules from that era are more likely to include commas and colons as just separators between elements where, in the case where you are doing minimal punctuation or records without punctuation, you end up with coding only with no punctuation included.

We've made the decision to keep using ISBD punctuation in records in our local database because the display in our PAC is ugly/confusing if punctuation is omitted.

That is understandable and is usually what drives the decision that libraries make about this issue.

For these punctuation records it will be interesting to study the Discovery display that OCLC makes (in Record Manager when editing use menu Record / View in WorldCat Discovery).

Just as is the case with local systems, there is a need to make changes to Discovery to better display records that lack punctuation. Of course, the German records have already been mentioned, which we have been loading for many years. They lack punctuation and they may be displayed less than perfectly in the Discovery. We have been working with our Discovery colleagues about the kinds of things that may need to be done in an environment where records lack punctuation.

It appears that these guidelines make the assumption that the default meaning of 245 subfield $b is "other title information", and that any programming to supply punctuation in a display will have to look at the content of the subfield $b to determine if it's starting out with a different mark of punctuation, to prevent an erroneous space-colon-space being added.

That's true, it will have to do that. We modeled, within PCC, the practice of relocating the equal sign and the semi-colon for the two specialized cases following the practice that had already been set in place by the Germans. It didn't make sense to us to somehow handle it differently. Along the way, there had been the question about if we are retaining two pieces of punctuation shouldn't we maybe retain all three. It shouldn't be that problematic for a program to look at the content that it is displaying and see that there is already an equal sign here, as the first character in $b, therefore the inclusion of a colon preceding $b is not needed.

General questions

March 2, 2020

If a record has a series statement that pertains to a local collection (for example, indicating that the piece is in a local archive), should it be removed from the WorldCat record?

For example, "Making of Modern Law" as used by Gale. When I change these records to a 040 $e pn then I want to remove this because it doesn't apply to all other sources. Or HeinOnline collections, same situation.

This is an interesting area because so often headings for local collections are established as corporate bodies rather than series. We have certainly seen institutions handle it both ways. For the most part, I would look at it and say that it is purely local information that is not of widespread interest outside the institution, and it really should be treated as
local information and not included in the WorldCat record. It sounds like it would be okay to possibly remove it. If it was a high-profile collection, a special collection within some institution, it could presumably be retained, but I would question if it should be retained as series or if it should be put in as a 710 with the name of the collection with a $5. Another deciding factor could be is it incredibly rare material that is only contained in one institution, in which case it is certainly fine for that information to appear. But if it is held by hundreds of libraries, then it is probably not such a good idea to have it in the record. So, it's a judgment call.

If there is a provider-specific series in a record that is intended to be provider neutral and is available from all providers, then the one specific provider series is not applicable to all instances of that resource online and can be removed. In these cases, there are a lot of records that appear with an 856 link to that specific collection, which is appropriate.

Field 710 with the name of the provider, any 490/830 fields with a series that is based on the provider are not applicable in a provider-neutral record.

I have been attending the RDA new concepts series and there are yet guidelines on what MARC fields will be used for some of the new concepts like representative expressions. Is OCLC planning for these?

As soon as there are new fields and/or new subfields for some of these new concepts, we will implement them. We can't implement them until they are defined. There is a lot of discussion going on about how some of these new ideas will be represented in bibliographic records. That process is discussion papers or proposals go before the MARC Advisory Committee, and once they have approved it, then the Library of Congress issues a notice saying that these are now official. At that point, OCLC will work towards implementing the new fields but not before.

March 12, 2020

Is OCLC working with ExLibris to allow cataloging functions directly in WorldCat from the Alma interface? Can you do the same things as Connexion?

We do not have knowledge if that is happening or not, so we can't really comment. That would be a question to send to OCLC Customer Support, as a starting point.

April 2020: Updates to record validation and the OCLC MARC update

Topic-specific questions
April 7, 2020

Is there any place that shows differences between authorization levels?

There is a table in Bibliographic Formats and Standards, section 5.2, that outlines the different authorization levels and capabilities for Connexion. You can also find information on the support site here.

Why are so many validation errors allowed in group batch processing?

When working on a single record, it is relatively easy to see what the validation errors are in order to fix those. When we are talking about large numbers of records, there is no way for a cataloger to individually track each record that has a validation error. There are reports that give the validation errors after processing is completed in DataSync that the institution may choose to follow up on and correct the validation errors in the records in WorldCat.

How do staging records play into validation?

There is validation that goes into DataSync but it’s not a strict as online validation, so more validation errors may be present in new records that are added to WorldCat. Incoming records are validated during processing and if they have validation errors, there is a level assigned from 1-3, with 3 being the most severe and will prevent records from being added to WorldCat as fully indexed without the records being corrected. Level 1 errors are minor and generally very easy for users to correct if they encounter them when working with records.

Are these validation checks only run on single records when they are saved, or are they run on batch loads too?

There are validation checks that are run on single records on certain points in the workflow including when you try to replace a record. There are also validation checks on batch-loaded records as well, as described above.

I occasionally run into records that have validation problems already in WorldCat. Are records validated before batch loading?

Yes, but there are less serious validation errors that may make it into WorldCat in order to get the records added to WorldCat.

How is verifying MARC-8 characters related to this process?

Since we removed checking for validation for Unicode characters, allowing pretty much any characters to be used in authority records, we don’t provide feedback on when characters outside of the MARC 8 rule set are present. We rely on our users to enter the correct characters.
April 16, 2020

Is there a list of validation errors?

No. The variation in the possibilities of validation messages would be huge. It’s not possible to list all validation errors.

After successfully uploading individual Level I English language records of Hebrew print materials, after some time I find they have been changed to Level M. How is this possible?

An example would really be helpful but it’s possible that the record is being overlaid by the same institution when the record is being sent back through their DataSync project and overlaying itself. If you encounter this situation, report that to bibchange@oclc.org so we can look into it.

Are those potential corrections only up to the institution or does Quality Control get involved?

Metadata Quality gets involved when made aware of any quality issues, including validation errors. Report those to bibchange@oclc.org when you encounter them.

When authority records are added to WMS, are there any authority records being put to staging?

No. NACO authority records automatically get put into the queue to be sent to LC. Staging only applies to bibliographic records.

I frequently come across Encoding Level M records that are completely full and have always assumed that these are batch-loaded records. Since we use automated algorithms to find copy cataloging records, these M records are rejected. Is there a way that these records can be more correctly coded?

Any record that is coded M by definition has been batch-loaded. As you may or may not have heard, we are in the process of trying to slowly do away with the OCLC-defined Encoding levels, including I, K, and M, and making the transition to using only the MARC-defined codes which is blank and numeric codes. The thing about Encoding level M is that it represents two things that should not be combined into one code. It does mean that the record was batch-loaded, but it also used to mean that it was a minimal level record, but that is absolutely not true in many cases, they can be of any fullness. If your algorithms are rejecting M as minimal, that is something you should look into changing.

As part of our eventual transition of Encoding levels to using exclusively MARC 21 codes, we are going to try to assign more accurately encoding levels when we convert them from I, K, or M to numeric codes and make them more reflective of the bibliographic records themselves.

Our VAOH in June is going to be on the subject of encoding levels, so you might want to tune in.

I was hoping that there was a list with solutions. I received a 1e error in Connexion (might not be the exact error message since that happened last week).

You can send the OCLC number bibchange@oclc.org and we could investigate to see what is going on with the record.
I often see validation errors in the 084 field for other classification numbers. Figuring out what's wrong with them can be challenging if you don't use that classification scheme or if they have been incorrectly placed in the 084 field.

We would encourage you to report that to us and we will try and figure it out.

When I see "049 too short" I just delete the field, but I don't really understand what it is about.

The 049 code has a requirement of 4 digits. It isn’t retained in the WorldCat record, so if you delete it, it’s not hurting anything.

Are validation scans being continuously run, like DDR or other quality control routines?

Only on authority records as they come in from being distributed back from LC. We get a report of validation errors in NACO records and we correct those.

Can you go over again what happens when an authority record is submitted with non-allowed diacritics? In that scenario, how long does the authority record usually stay in distribution until it is corrected?

When a record is sent off to LC, we will get a report back, usually the next day, that the record is being rejected because of incorrect characters. Metadata Quality staff go through that report usually on a daily basis and correct the records. It may take several days for the record to go back through the distribution. Our turnaround for authority records that are stuck in distribution is pretty quick. If you find a record that has been in distribution for more than a week, send an email to authfile@oclc.org.

Non authorized MARC fields and subfields that have not been implemented by LC should not be used in authority records. That can also cause records to get stuck in distribution.

In our weekly OCLC error report for batch-loaded records, sometimes we see "Sparse Check failed. Sparse record" What does it usually mean?

Please see Sparse records information at this link.

Are local holdings records validated?

Yes, and for the most part they follow the MARC 21 holdings format.

How is it that some records get into the system with validation errors and then when we try to set holdings, we are unable because of these errors. Are these not caught by machine?

There are various validation levels for records that come in through DataSync and some of the less severe errors are let in that can be corrected either by the institution or by OCLC. Batch-loaded records may have validation errors that would not occur if a record were added online.

I use LC classification, but often get validation errors for 084 fields. I don’t know how to correct them so I can replace the WorldCat record.

If you are not sure how to correct an error, you can report those to bibchange@oclc.org and we will correct if for you. It’s also useful to have patterns of errors reported to us so that we can relay the information back to the contributing institution to have them correct their records in future loads. We are also working on a solution to prevent validation
Can the last validation error mentioned by Bryan be fixed? I have produced NACO authority records only to find out weeks later that the record was never produced because it contained a pre-composed character with a diacritic, which I didn't know about when I produced the record. Can't validation catch this?

We've been looking not how records with precomposed characters get into the process and somehow are exported with the pre-composed characters. It seems that there may be something within the Client that de-composes the characters upon display, but still sends them to LC as pre-composed. Nothing is transforming them as they should be and we are looking into it to try and figure out where it's not going as expected.

When getting an invalid character message, it is sometimes very difficult to determine exactly where/what the error is, it's necessary to count the characters. Would it be possible for the error message to display what the invalid character is?

Unfortunately, no. Validation can't be that specific because of the way that the messages are used from templates. They can't be as specific as “this particular character” is incorrect.

(OCoLC)01837402 has multiple 505 fields with pagination - I just have not seen multiple 505 fields created in that fashion. I suppose since this is variable, there isn't any kind of check on this type of field?

These field transferred during a merge. The record will be corrected.

Another instance I have found when no error message is displayed, but the record will not validate, is when 336 $c is used erroneously instead of $2. Why is this?

That is a mystery, so please report those in the future if you encounter it again to bibchange@oclc.org. All of the validation errors are created by templates and they have to be manually formulated by relationships that are checked by validation. So, if there is an error that doesn't generate an error message, it must be that we missed that.

Sometimes I want to change a monographic record my library created and my library is the only holder to a continuing resources record, but after I make all the changes it says I am not authorized to change the Type and I lose all my edits as well as the saved record if I saved it. Is there a way to avoid losing the edits?

If you are the only holding library, you should be able to make changes to Type and BLvl. Report that problem to bibchange@oclc.org to be investigated if you encounter this problem again.

How do we deal with these validation errors that are outside our knowledge areas? e.g. Dewey, etc.

You can report these to bibchange@oclc.org.

I thought you used to be able to catch precomposed characters in Connexion with Edit -> MARC-8 Characters -> Verify, but it hasn't worked for me lately -- did this change?

This did change when OCLC implemented Unicode, so now all characters are valid within Connexion. So, this particular verification is no longer able to be used.

Why not just switch the Marc-8 characters to Unicode in authority records?

LC's system does not yet accept characters that are outside of MARC-8.
Is there a way to find out the validation rules for local holdings records?

The use of the local holdings record is described here.

Are records displayed in WorldCat.org composed from multiple records and is there a way to find out what those are? For example, this record has a content note from a different book, but I don't see it in the OCLC records. Why?

In WorldCat.org fields can be populated into a representative record from other bibliographic records or imported from a 3rd party provider. These include summary notes, abstracts, and contents notes. Errors should be reported to bibchange@oclc.org and if the data is not from a bib record, we will forward the request on to OCLC Customer Support to be removed.

Also related to WorldCat.org -- A coworker told me that holdings used to appear as soon as he added the library's holdings symbol via Connexion Client but now that does not happen. Was there a change?

Holdings may not show up right away in WorldCat.org depending on browser settings. It is possible a cached version of the page is being displayed, to see the immediate change in holdings you may need to adjust your browser settings.

Other reasons they might not appear, the member has a cataloging subscription, but no subscription to WorldCat Discovery/FirstSearch. You need to have both subscriptions to see your WorldCat holdings in WorldCat.org.

Here is a link to our Help documentation on Why aren't my library's holdings displaying in WorldCat.org?.

May 2020: The mysterious 3xx fields

Topic-specific questions

Is there a list of the terms used in 34x fields available outside the RDA Toolkit?

BFAS 3xx fields page points to the appropriate vocabulary to use. You can also find the list of terms for these fields by going to the RDA registry's RDA value vocabularies.

Which fields are valid for both bibliographic and authority records?

Fields that are work and expression based would be valid for both. These would include: 046, 336, 348, 370, 377, 380, 381, 382, 383, 384, 385, 386, and 388.

A full list of 3xx fields for authority records can be found in the MARC 21 Authority Format, Headings General Information page. For a list of all of the fields that can be used in bibliographic records, please see BFAS, 3xx Fields.

Are there any examples of the 386 field being used in the public catalog interface? Anyone doing this? How did you choose to set up indexing, display, etc.?

Many of these fields and subfields were not intended for public display on their own. The intent of many of these fields is to enable a local system to facet things in ways not done before, such as identifying specific format. Currently, the OCLC Discovery interface does not display these fields. However, Valdosta State University
libraries use fields 385 and 386 in their public display. If you go to their catalog, you can see how they are used in display and faceting.

If I am cataloging a collection of letters and papers belonging to a single person would I include a 381 field with the persons last name or is that necessary?

In this case, you would have a subject access point for the person’s name instead of using field 381. However, use of field 381 is geared more toward differentiating one work or expression from another work or expression. For example, two different motion pictures released in the same year with the same title but with different directors.

Harlow (Motion picture : 1965 : Douglas)

Harlow (Motion picture : 1965 : Segal)

Can you expand why OCLC decided to use separate fields for each unique term even if the terms come from the same vocabulary?

While it is not required the you use different fields when the terms are from the same vocabulary, the thought was that it might be easier to facilitate things like faceting when searching if separate fields were used. Both OLAC and MLA best practices allow you to use a single field with separate subfields when the terms are taken from the same vocabulary or from no vocabulary at all. While current best practices allow terms from the same vocabulary to be added to the same field with separate subfields, this may not be the best solution when using subfields $0 and $1, which would be used in transforming data from MARC to BIBFRAME or a linked data environment.

Are the 3xx fields indexed in WorldCat?

Yes, you can find a complete list of the 3xx fields that are indexed in the OCLC Help site under Searching WorldCat Indexes, Fields and subfields, 3xx fields.

Does OCLC plan to have validation when subfield $2 codes are present in the 3xx fields?

Yes, the subfield $2 codes are already validated.

Is it recommended to fully code both 007 and 3XX fields despite their great overlap in content?

OCLC recommends that, for the time being, libraries continue coding field 007 while adding any appropriate 3xx fields. OCLC continues to participate in the current discussion about using 007 and 3xx fields. A lot of existing local systems were built to use the 007 field for faceting and differentiating one resource from another and not all of them have adjusted to using the newer 3xx fields. In WorldCat, OCLC uses both field 007 and 3xx fields to determine material type, while the Library of Congress has moved to using 3xx fields instead of field 007 and various fixed fields when converting BIBFRAME to MARC.

More of a discussion prompt than a question... I recently finished a collection-level record for a 163-linear foot collection of many material types. When I looked at the finished record in WorldCat, it picked out the 346 values for display, making it look like video formats were the only material types included even though 15/16 series are not video. I guess that what’s we want??? Our archivist thought that was so misleading that I went back and deleted the field at her request. For those viewing chat text, details are...

Displayed material types: Videorecording (vid); U-matic (umc); Videocassette (vca); VHS tape (vhs)
346 ǂ3 Series 9 moving image material: ǂa 8 mm ǂa Betacam ǂa Hi-8 mm ǂa U-matic ǂa VHS ǂ2 rdavf

300 (101 boxes, 15 oversize folders, 2 binders, 102 books, 5 items, 2 oversize containers, 1 drawer of notated music, 20 audio discs, 150 audiotape reels, 772 audiocassettes, 74 film reels, 231 video cassettes, 254 video cartridges)

There is a lot that goes into displaying the icons in WorldCat. When your library sees a specific display that is misleading, please send the OCLC number to askqc@oclc.org and we will look into it. The generation of Material Types from the data in a bibliographic record goes back a long time, significantly predating the definition of many of the MARC 3XX fields, including 346 (defined in 2011) as well as the creation of the RDA Registry that tries to codify many of the controlled vocabularies we now rely upon. Many of those newer 3XX fields are certainly taken into account in the formulation of Material Types; a few of those of more recent vintage are slated to be taken into account the next time we are able to make changes to WorldCat indexing.

For better or worse, not every possible kind of video or audio format generates its own specific Material Type, although all of the most common ones do (in video, for example, VHS, DVD, U-matic, Beta, and Blu-ray are among those that generate their own MT). The more obscure video recording formats (many of which are documented in MARC Video 007/04 or in BFAS Video 007 subfield $e), such as Type C, EIAJ, Betacam, M-II, 8 mm, and Hi-8 mm, instead generate a more general MT as appropriate. There were simply not enough WorldCat records representing some of these formats to justify their own Material Type. This explains why the record in question registered only VHS and U-matic as specific MTs; the other video formats listed in the 346 essentially roll up to the more general "Videorecording."

In a vast archival collection such as #1149392345, which includes numerous different kinds of print and nonprint media, there could have been literally dozens of Material Types represented, had all of them been accounted for in the bibliographic record. Coded Material Types may derive from various fixed field elements, 007s, 300s, 33Xs, 34Xs, 856s, and elsewhere.

On slide 23, you show the key of music as D minor. To musicians, a lower-case letter shows a minor key: d minor; upper case is for major key: D Major. Does MLA have anything to say about this?

The field in question does not have a controlled vocabulary and traditionally, the key was always capitalized (i.e. D) and the mode (i.e. Major or Minor) was spelled out. The best practice is to capitalize the key and spell out the mode. This partially has to do with indexing and other software like that. Rebecca Belford in chat mentioned that Capitalizing key in 384 also matches the AAP format for $r, making copy-paste or theoretical machine generation easier.

We have a recording that includes a CD, a 12 inch 33 1/3 rpm disc and 7-inch 45 rpm disc. When recording the 3xx fields, is it best to create separate 300 fields first and then how do you organize all of the 33x and 34x fields? There would seem to be a lengthy number of fields and some might overlap. Any suggestions?

RDA allows catalogers to use separate 300 fields or string them together in a single 300 field. A suggestion would be to organize the 3xx fields in the same order in which they appear in the 300 field(s) and use subfield $3 to identify which piece that particular 3xx field refers to.

Should we be using tag 385 instead of a 521 note for intended audience?

For now, OCLC encourages the use of both fields during this transitional time. While field 521 is intended for display,
field 385 is not intended for display.

If many fields are not intended for display, why then do we use textual controlled vocabulary, which is prone to typos?

The answer is that it is more practical. Both textual controlled vocabulary and codes are prone to typos. While RDA sometimes gives one preferred vocabulary to use, for many fields, there are a number of vocabularies which would require the creation of a corresponding set of codes for every single vocabulary.

How granular should we be? For a recording with both stereo and mono tracks, or for a SACD (usually coded both stereo and surround), is it desirable to enter these terms in the same 344 field or in separate ones?

The extent of the granularity and how much information you want to facet is up to you. In field 344, it would not be uncommon for more than one term from the same vocabulary to apply to the same resource. As mentioned in the presentation, the MLA and OLAC best practices both allow catalogers to use multiple subfields in the same field when using the same vocabulary or no vocabulary at all. While OCLC prefers the use of separate fields for each term, catalogers may choose what works best for them.

Does OCLC have any plans to partner with anyone to retroactively add some of these fields to legacy records? I know that MLA is working on a toolkit that will add some of these fields to legacy records but not all. Are there other projects underway from other communities?

Gary Strawn from Northwestern University has created a toolkit that will create certain 3xx fields based on information elsewhere in the bibliographic record. OCLC has looked at applying Gary Strawn’s software to WorldCat to apply some of these 3xx fields. Although no decision has been made one way or another, we are interested in any ideas about how to retrospectively add 3xx fields.

Was the GMD deprecated because it was redundant with the 3XX fields?

No. The GMD was always problematic because it was one dimensional and trying to do many things with one piece of information. The International Standard Bibliographic Description (ISBD) tried to facet out what GMDs were trying to say in the sense of content, medium, and carrier. This was the origin of the 33x fields. The 33x fields were designed to replace the GMD during the transition from AACR2 to RDA.

Is it possible to use subfield $8 in addition to or in place of subfield $3 to connect multiple kinds of 3XX fields?

If your system is set up to deal with subfield $8, then you are welcome to use it. There are not many systems that use this subfield. While you may use it, OCLC does not suggest that you replace using subfield $3 with subfield $8 as subfield $3 is displayable and human readable and the subfield $8 is not.

I notice on some slides that this presentation includes the RDA instruction number. Will future presentations make use of the citation numbering used in the RDA Beta Toolkit? I am trying to learn the Beta Toolkit and use of the citation numbering.

Currently, the RDA Beta Toolkit is not in its completed form, as the Beta site itself notes: “... the functions and content of the site are still under development. The RDA Steering Committee has not yet authorized the beta site for use in cataloging work. The beta site will become the authorized RDA Toolkit on December 15, 2020.” Even at that time in late 2020, both the current version of the RDA Toolkit and the Beta version will be available. A year-long Countdown Clock will begin sometime in 2021, at a time yet to be determined by the full agreement of the RDA
Board, RDA Steering Committee, and the publishers of the RDA Toolkit. We will have to see how things develop, but we would imagine that OCLC presentations created during the period when both RDA versions are available will reference both the original RDA instruction numbers and the new numbers and that we will switch over entirely when the original RDA is decommissioned. Until December 15, 2020, we will continue to use only the original RDA instruction numbering.

Is OCLC’s recommendation to put subfield $3 at the end rather than the beginning of a field (as in the example for the audio recording that also had a guidebook, and some repeated 33x)?

MARC 21 illustrates, with some examples, subfield $3 at the end of the 33x fields, so when OCLC implemented these fields a few years back there was a conversation about what the recommended practice for placement of subfield $3 should be. Catalogers are used to placing subfield $3 up front with other fields for display purposes. When looking at the 33x fields, though, the thought was that since they were primarily there for retrieval and indexing purposes and not for display, then subfield $3 is nothing more than a control subfield and so should be listed after subfield $2 at the end of the field. If your library decides to display these fields, you may move the subfield $3 to the front of the field for use in your local system.

What is the relationship between field 380 (Form of Work) and 655 (Genre/Form)? What are the best practices for using one field vs. another to express the form of a resource?

Adam Schiff answered that typically, field 380 would be used to record a generic genre/form term, usually the one that would be used in a qualifier in the access point if it were needed. Field 655 would have the specific genre/form terms. For example:

380 Motion pictures $2 lcgt

but

655 Comedy films. $2 lcgt

380 Fiction $2 lcgt

but

655 Novels. $2 lcgt

655 Detective and mystery fiction. $2lcgt

Jay Weitz commented that OCLC doesn’t offer specific guidance about these two fields but, generally speaking, these two fields have different purposes, and, in WorldCat, they appear in different indexes.

I would love to see a controlled vocabulary for the Accessibility field, and echo Bob’s question about where they’re indexed so I could see examples of how they’re being used ‘in the wild’.

Honor Moody mentioned that there is not a universal accessibility field, but that the W3C has an accessibility schema that can be considered a controlled vocabulary.

What would you use for the 380 for Made-for-TV movies? ‡2lcgt and Fiction television programs. ‡2lcgt in the
655 fields for 24. Redemption?

Adam Schiff said that OLAC best practice says to give "Television programs" in field 380 and the specific terms in field 655.

Would you include 380 for nonfiction too or let 655 handle it? For example, Personal narratives, Guidebooks?

Adam Schiff said that his library typically doesn’t bother to give a 380 field at all, since the 655 field can be used for all the appropriate genre/form terms. Their system indexes both field 380 and field 655 as genre/form terms. It really depends on the system your library is using and how it’s configured.

On slide 34, "polychrome" is synonymous with "color" so are we saying that we can just put in a different word if we don't really like the RDA term?

Adam Schiff mentioned that different terms from different sources for the same concept can always be recorded in most of these fields. Kelley McGrath added that there is a history behind the polychrome problem and the way RSC wanted to define color vs. the way it applies to tinted and toned film, which is why there’s an alternative to use a substitute vocabulary.

What is everyone using in field 347 for an electronic book?

If it is a text file, you would use the phrase “text file in field 347. For example:

347 text file $2 rdaft

Does OCLC have a stance on privacy concerns related to recording demographic characteristics (especially gender) of creators/contributors in 386?

This is a very sensitive topic, and while OCLC doesn’t have a stance on this, the general consensus is that just because you have the data doesn’t mean that you should record the data, especially with sensitive information. So, the cataloger should be very deliberate when making the decision to include sensitive information or not.

Is there a list of official abbreviations to be used in the $2? I know they’re based on the titles of various RDA sections, but the longer the section title, the harder it is to guess what the abbreviation should be.

The codes in subfield $2 are not abbreviations but rather codes. All of the codes used in subfield $2 can be found in the appropriate MARC code list on the MARC website.

Can you please expand on using the 300 fields for born-digital materials that come in via cloud transfer, with no container?

They would be treated as any other online resource.

Depending on the format of the eBook, is it really a "text" file? like MOBI or EPUB?

Adam Schiff stated that he assumed that text file means something you can read with your eyes. Bryan Baldus added that the RDA registry says that a text file is a file type for storing electronically recorded textual content.
Encoding the 3xx fields can take a lot of time, is OCLC planning to provide some help to streamline the creation of these fields?

Yes, there are some tools out there that may help you. Gary Strawn from Northwestern University has created a toolkit that will create certain 3xx fields based on information elsewhere in the bibliographic record. There may also be other macros out there to assist you in creating 3xx fields. Robert Bremer and Jay Weitz have talked about going back through WorldCat to try to retrospectively create various 3xx fields, however, this has not been done yet.

Is there any thought for OCLC to creating separate indexes to better search for specific characteristics? For example, limiting a search to just 385 or 386 instead of all other 3XX fields?

The entity attributes index was created a while ago, during the early development of many of these fields and before some of the fields existed. At that time OCLC had very little idea how they would be used, if they would be used, and what kind of vocabularies would be used in some of these fields. Because of this, the decision was made to utilize the entities attributes in the short term, with the intention to possibly creating more specific indexes for individual fields as needed.

Is it still best practice to code 336 still image for books that contain a lot of images?

Yes.

Is OCLC retrospectively separating 34X fields with multiple subfields into separate 34X fields for each subfield?

OCLC, currently, does not have any automated process doing this but it makes sense that we ought to do it. Terms that have been supplied in some of these fields that match up to pre-RDA controlled vocabulary could have subfield $2 added with the appropriate codes to clean up the records.

I heard a partner of OCLC has some of the virus, to see how it reacts with various library materials. How is that going?

OCLC has partnered with Battelle to discover how the COVID-19 virus works on various library materials. Metadata Quality staff are not involved with this project so we cannot answer specifics about it. For details about the project, please see Reopening Archives, Libraries, and Museums (REALM) Information Hub.

Jumping back to the earlier 347 eBook question regarding text files, what happens for a children's book where the text is minimal but it's heavily illustrated?

For an electronic book version of a children's book, if you choose to use field 347, you would use “text file” and “image file” to bring out both the text and illustrations.

I also am wondering about controlling headings in these fields. Will there be that functionality at some point like there is in 6XX and other access point fields?

While we think this is a great idea and would be extremely useful, OCLC has not yet looked into this and what would be needed to make this happen.

Adam Schiff said that regarding Temporal Terms Code List, the ALCTS/CaMMS/SAC Subcommittee on Faceted Vocabularies is considering creating a controlled vocabulary for chronological headings that could be used in the 388 field. Stay tuned. He also mentioned that the SAC Subcommittee of Faceted Vocabularies will be issuing best
practices for recording 046, 370, 385, 386 and 388. The one for 046/388 is nearly complete and hopefully will be published by SAC later this spring or summer.

For adding 3xx fields that are standard for print or DVDs or audiobooks when doing original cataloging we put these in word documents and copy and paste them into OCLC when creating the record.

Yes, this is a great workaround for adding the same information to the records you are creating. Another option would be to use constant data.

Do you recommend putting all subfields in separate 3XX fields even if the $2 would be the same, such as 340 $a, $c, and $e that all use the rdamat vocabulary?

Yes, this is OCLC’s recommendation. This is partially due to the possibility of including subfield $0 and subfield $1 to these fields, in the future. The subfield $0 and subfield $1 would be associated with an individual term or phrase in a subfield and would be used in transforming data from MARC to BIBFRAME or a linked data environment. So, putting each subfield in a separate field would facilitate this.

Wouldn't creator characteristics be better in authority records and not entered in individual bib records? Or do we enter in bibs because systems can't handle it from authority records?

Yes, putting these personal attributes for a creator are appropriate in the authority record. It is your choice whether you put them in both places or not and may have to do with your local system capabilities. As we look forward to a world of linked data, new best practices may emerge about the optimal place to record such information. This is likely to be an ongoing conversation within the cataloging community.

### June 2020: Updates on OCLC encoding levels

#### Topic-specific questions

If we are upgrading a K, M or J level record with item in hand can we change the encoding level to blank?

Yes.

Could you talk about who is contributing most of the ELvl M records? Are they mostly vendors?

The majority of M level records come from member libraries. There is a certain percentage that come from vendors because the majority of vendor records that are contributed to WorldCat do come in through Data Sync and so are assigned Encoding Level M. But there are huge numbers, much more than the vendor records, that come in from member libraries and those are all assigned Encoding Level M when they go through the regular Data Sync process.

For the documentation, it may be useful to have a page somewhere that includes information on the alphabetic codes, for legacy purposes. Even if they’re removed from OCLC, they may not be removed from local catalogs.

We’ve talked about the need to incorporate history into Bibliographic Formats and Standards for reasons just like this. When you find something on a record that you have in your database that is no longer current coding you kind of want to know the history behind that. We tend to look at MARC21 for some of that information but in the case of something like Encoding Levels I, M, K that were OCLC-specific that information is not going to be there. So yes, it
makes a lot of sense that we should have some kind of history section for this information.

**Have you also been consulting vendors of library systems who will need to incorporate this?**

No, we haven’t done that yet and the main reason we don’t feel there is a big need to do this is because all these new Encoding Levels that members libraries will be using are all already valid as part of MARC21 and have been for years. If library system vendors have been loading records from the Library of Congress, or records from the Library of Congress that you’ve obtained from OCLC they’re already familiar with loading these numeric Encoding Levels.

**How will this be disseminated to cataloguers all over the OCLC world, who haven't attended this session?**

We’ve discussed the need for additional training that we would like to put out there because we know not everybody necessarily tunes into these sessions. But yes, we will get the word out in advance, so people are used to using these new Encoding Levels long before we make them invalid.

We put these out in the Release Notes that came out for the April install, so there was information there. The recordings of these sessions will be available, as are all these Office Hour sessions, and we will start promoting recordings and the information in a big way as we get closer to making other changes. But for right now, we just want people to get a chance to take a look and think about using them. There isn’t a requirement for people to switch right now unless they want to.

**What codes should we use now if we are entering a full-level original record? We were recently instructed to use ELvl "I" for everything. What codes should we use now if we are entering a full-level original record? We were recently instructed to use ELvl "I" for everything.**

You can continue using Encoding Level I, but at this point you can also enter a Full level record as Encoding Level "blank." They are essentially equivalent and “blank” will become preferred as we go forward.

**Is there a coordination underway with local databases to assess the meaning of "M" in individual records?**

M really has no meaning in terms of Encoding Level. It represents a status of the record in OCLC’s system. If you’ve taken a record from OCLC and put it in your local database and retained Encoding Level M it really doesn’t have too much meaning, there. It may be that locally you may want the Encoding Levels that we would eventually change these records to but it’s kind of hard to say. We also are considering what the impact is in changing records in the database. Some libraries would subscribe to Collection Manager and receive all of those updates, some libraries, of course, wouldn’t want the volume of updates coming through. But if you did incorporate changed records received through Collection Manager you could potentially update your own database to get rid of the OCLC-defined Encoding Levels.

**What is the difference between code 3 and 7?**

Code 3 is for abbreviated level, which means less than minimal and 7 is minimal level. So, 7 is equivalent to K. We do have, in one of the first chapters in Bibliographic Formats and Standards, some information in a chart there about abbreviated level records and what one would normally use when encoding a record as level 3.

**Is there training documentation available yet?**

Not specifically designed as training, but I would direct people to have a look at the Encoding Level page in
Bibliographic Formats and Standards, it pretty well explains the situation, especially when you get to the OCLC-defined codes that are at the bottom of the page. It indicates what’s happening with them and what code would be preferred in place of level I or level K.

Why is it important to know if a record was batch loaded? For example, couldn’t encoding levels 1 and 2 be used based on the completeness of the record?

For the first part of the question, we received feedback that there was interest in knowing that a record had arrived in the database via Batchload; because it is something that we could just get rid of entirely. But a lot of libraries were interested in having that information at hand. It is useful, in some respects, especially when you are looking at records that look very similar and you’re considering are they duplicates or not. If somebody intentionally put it in, it may be that they have a different version of a record. If it only arrived by machine processing it could be that it wasn’t detected as a duplicate. So, in those terms, knowing that it came in through Batchload is a useful kind of thing.

2nd part of question: In a sense, maybe they could, but it’s not as if it’s data that hasn’t been examined, which is part of the definition of 1 & 2, in terms of the record that’s in OCLC. A cataloger really did look at the original item and perhaps supplied Encoding Level “blank” and then sent it to us. Changing that “blank” then to be a 1 doesn’t seem to be the right thing to do at this point. It would be better to keep the “blank” intact, because somebody did have the item in hand when they created the record, and then store the indication that the record arrived via Batchload in another field.

Does encoding level 8 become an M in a tape load? When doing a merge, is there any way to know this?

Encoding Level 8 becomes M depending on, in a Batchload situation, what library it comes from. From member libraries it does indeed become an M; from the Library of Congress, or other national libraries where we get CIP records those remain Encoding Level 8. It does depend on the source of the record right now. In the “future world” we will be retaining 8 from whatever source we get it from, but that’s not implemented yet.

And “When doing a merge is there any way to know this?” Not really. If it’s an M, right now, if we are doing a merge, or if one of the Member Merge participants is doing a merge, they need to just examine the fullness of the record to figure out when it came in, and what it was when it came to us. It’s a little bit of guesswork.

Didn't "K" use to mean "less than full" rather than its present "minimal"? [https://web.archive.org/web/20050811235459/http://www.oclc.org/bibformats/en/fixedfield/elvl.shtm]: "Less-than-full input by OCLC participants. A record that conforms to OCLC's level K input standard. The level K input standard represents less-than-full cataloging."

I don’t recall that it was ever defined as less-than-full, but I hesitate to disagree with Walter Nickeson. I think it’s always been defined as minimal, as far as I can remember.

Will there be improvements to data sync so that incoming batchloaded level M records will match on existing ones and not create so many duplicates?

That’s a general question about Data Sync and the matching that goes on in that process. We do look at that constantly to see what it should do that it’s perhaps not doing at this point. Once a duplicate is added to the database it is subject to DDR processing which compares records in a somewhat different fashion than is the case with Data Sync. We have two chances to catch duplicates as they come in.
The problem with so many duplicates is that there is something in the record that prevented it from matching. And the very kinds of things that you could look at a record and say: Well, this difference doesn’t really matter – in one case is the same kind of difference that in another case indicates that there really are two versions of some resource. It’s a very fine line to get things to match correctly, but not necessarily match things that really are different. So, again, we are always looking for improvements to that process.

The question about M and local databases is meant to be the other way around: if the local databases have specific ELvl, but OCLC only has the M, could the OCLC level be changed to match what the local databases have.

In that case we probably still need to do an assessment based on the data that is in WorldCat because, let’s say we have an Encoding Level M record and it’s pretty skimpy but that same record exists in some local database and is in that local database upgraded to a level “blank” it could be that additional fields have been added locally as well, to make it a full and complete record. Well, in WorldCat we might still have something that really ought to be considered minimal level. It is all based on what we have in WorldCat in terms of making an assessment and figuring out what the Encoding Level should be.

Why were full records, cataloged in OCLC, encoded 'I', changed to M? I have seen this with a large number of full records in OCLC.

That happens in a specific instance of batchloading when a library who has entered the records in WorldCat with their OCLC Symbol as the creator then also sends us those same records via Batch (Data Sync). One of the options in the Batchload profile (the Data Sync profile) is to check whether they want their own records replaced. If they have that checked, that means if no one else has modified the record the records from that library will replace their own records in WorldCat and change the Encoding Level as a result because all of the Data Sync loaded records are Encoding Level M. So that’s why that happens.

And will the BFAS Quality Assurance instructions be updated as well?

Correct, there are some changes that need to be made to Chapter 5 that discusses quality assurance to bring them into step with what the current situation is.

Will you share the parameters you use to convert the level Ms so that we may use them in our local databases?

We’d be happy to share them. They aren’t ready yet. They have been drafted but we still need to test them and make sure they are complete and cover what we need them to cover. Once we have them ready, and perhaps tested them a little bit, we would be happy to share them. I don’t know when that will be. I suspect it might be next year some time.

So, is the idea to eventually get rid of all of the OCLC Marc Encoding Levels of I, K, M, and J?

You got it. That’s the major point to take away today.

Do you want us to start using "blank" when we lock and replace a level "3" record instead of an I?

You are most welcome to do that. It really is up to you now, whether you want to use I or “blank,” those two levels are equivalent, so why don’t you start experimenting with “blank” because you are able to do that now when you are working online.
Will you receive a validation error if you accidently use blank and d vs. blank and c, and there isn't an 042 field in the record?

I do not believe that there is any relationship in Validation between those elements and it kind of makes sense that there isn’t because you could have input copy in the past that would have required use of Source (008/39) “c” in combination with an Encoding Level like “blank.” In other words, we have combinations of Source ”c” with Encoding Level I as it is now. So, Source “c” is not tied to Field 042 at all. And, of course, Source “c” does get misused and it does cause things to get incorrectly indexed. So, we’re aware that there’s an issue there.

Looking through BFAS, I realize I’d forgotten than ELvl 5 even exists. Is that the ELvl some of LC's overseas offices use for their brief bibs before they come back and finish them later? Do you see other libraries using 5 often?

Other libraries will have most likely not need to use 5. We will not have anything in place that says you cannot use Encoding Level 5. But if you were to use Encoding Level 5 the expectation is that you’re going to come back at some point and finish that record off and upgrade it to something like “blank” or even 7. Other libraries, of course, could come in on that record in the meantime and change Encoding Level 5 to be something else. But, for the most part, as it is now, libraries generally complete the record input it into the database.

I have tried to edit a level blank record and was not able to replace the record. Will there still be level blank records OCLC members cannot update? Are there still fields in level blank records we cannot change?

You should be able to edit a level "blank" record in the same way that you would have edited an I level record in the past. PCC (Program for Cooperative Cataloging) records are still exempt if you are not a PCC member, just as they have always been.

In terms of elements you cannot touch, system-supplied kinds of elements in the record that you cannot change (etc.). None of that has changed.

One of the presentations that we did a few months ago in these Office Hours included a large section about what you can and cannot edit in PCC records. That was the February session [Best practices for enriching WorldCat bibliographic records]. You may want to take a look back at that to get instructions and then see the references as to where in Bibliographic Formats and Standards we outline what may and may not be edited in PCC records.

L has already gone away.

That’s true, we eliminated Encoding Level L a few years ago. That was considered Full level from a Batchload and it just seemed simpler to have one level from Batchload which is M.

Does that status as growing, for ‘M,’ mean that records have been downgrade, or that records are being added as ‘M’?

It means that records are continually being added as Encoding Level M.

How soon do you expect to have additional training about using the numeric values?

We don’t have a timeline in place, but we will start working on that in the second half of this calendar year. We’ll release it and let you have it as soon we have something. We’ll certainly have new training in place before we make any massive changes within the current WorldCat database.
Does it mean, when we create original cataloging records now, we will start to code blank in ELvl? For I level?

Yes, that would be a good thing to go ahead and start doing. “Blank” is the equivalent of I so you can switch from using an Encoding Level I for a Full level record that is brand new and use “blank” instead.

When enhancing or adjusting an I-level WorldCat record, is it acceptable to recode ELvl I as blank?

Yes, that would be a great thing to do. When you are upgrading, or otherwise editing a WorldCat record and it’s coded either I or K and you want to change it, or if it’s coded M and you want to change it to “blank” or 7, please feel free to go ahead and do so.

Do we think the continual growth of M is caused in some way by vendor created records?

No, it’s every record that comes in via Batchload (Data Sync) or via the WCIRU process – we have a lot of these batch processes – that is added to WorldCat is made into Encoding Level M. I shouldn’t say every, I should say most records, the vast majority of records coming in that way. So that’s why. Many, many libraries send us their records via those batch processes, particularly via Data Sync. So, it’s not just vendors. Vendors are only a small percentage of that, maybe between 5 and 10 percent of the records that we add through batch processes.

Those of us that send records for batch processing, these records will be set to BLANK if they come that way?

Yes, in the future we would retain the Encoding Levels as they come in. If we received a record that was Encoding Level “blank” it would end up added to the database that way if it didn’t match another record. That’s the significant difference between where we would like to be in the future vs where everything is arbitrarily set to Encoding Level M.

Is blank the pipe key?

No, “blank” is the space bar when you are typing on a keyboard.

M may also be from data sync ingests?

That’s correct that’s what M is from.

Could you explain some of the differences between MARC21 and OCLC-MARC?

We have been trying, quite purposefully, over the past few years, to eliminate many of the differences between MARC21 itself and OCLC-MARC which is OCLC’s implementation of MARC. There are some things that we have not yet eliminated, including the Encoding levels. If you go to Bibliographic Formats and Standards, the contents page of BFAS, there’s a document linked from there that spells out most of the remaining differences between MARC21 and OCLC-MARC. One of the big ones is the use of the local field 538 which OCLC defined as, for various reasons, mostly having to do with display in previous platforms for the database, the use of 539 instead of subfield $7 in Field 533. That’s one that always sticks out in my mind, but there are a few others as well.

Are there requirements for particular fields at a given encoding level?

If we want to back up a few slides to the display of the Input Standards, the one that has the 3 red boxes on it, the Input Standards for the field as a whole is given at the top of that display. [Slide 18, BFAS Documentation Changes]. Right where it says Input Standards below that is the field level input standards Full vs Minimal, so in this case, which is actually Field 300 it’s Required if applicable for Full level records and for Minimal level records. That’s the Input
Standard at the field level, then, of course, we also have the Input standards at the subfield level. So, the expectation is, if you are going to use “blank” for Full then you would be following the Input Standards on each of these pages. If the field is Required if applicable, and it does apply to what you are cataloging, or it’s mandatory for Full level, you would need to input that.

Is core level 4 obsolete? What will happen to level 4?

Encoding Level 4 really is obsolete; it was tied to the previous Core level record standards and those standards are now obsolete. There isn’t a scenario where you would use Encoding Level 4 on current cataloging. It's still valid in the system, because records do exist in WorldCat with Encoding Level 4, but it’s not as if that number should be growing.

So, when M is going to disappear?

I would say Encoding Level M would disappear at the point that we have changed Batchload processing so that we’re no longer creating new Encoding Level M records and we have converted the last one. We really can’t take anything out of our Validation rules until all instances of that particular code have been removed from the database, otherwise if you go to use that record for copy cataloging and you decide to validate the record you will get an error message that says Encoding level isn’t valid. So, of course, we should fix that for you upfront so that that doesn’t happen.

Encoding Level M, because it is the largest group will probably take the longest to eliminate from the database. It’s probably several years out.

What is the best encoding level for analytic records?

It all comes back to how much detail you are including in these records. If it’s an analytic record that is fairly brief, then maybe you would end up with Encoding Level of 7. But an in-analytic, an article that appears in a journal, doesn’t have a whole lot of detail anyhow so it may qualify to meet the full level record standard anyway. In that case it would be Encoding Level “blank.”

Will LOC records continue to be clearly identified, as I think they, national libraries and PCC were the only ones that had previously been able to use BLANK.

Yes, Library of Congress records of course indicate that they have come from the Library of Congress in Field 040, and the same is true of other national libraries. But the authentication code in Field 042 is usually pretty important in identifying that a record is PCC (Program for Cooperative Cataloging) and meets certain other standards in terms of authority control.

Is there any way to have the encoding level assigned automatically applied based on an examination of the record against a set of criteria?

We’re sort of hoping that’s the case, because the approach to dealing with eliminating Encoding Level M is to examine the record and see if it looks fairly complete. Of course, in doing that you have to take into consideration a lot of different factors. The way a manuscript letter may be cataloged ends up with a description that’s fairly brief even though it would be considered Full vs a published book for example. You have to consider the coding in the record, what kind of material it is, in order to assess whether it appears to be complete.
If all the differences between OCLC coding and MARC 21 are removed, will BFAS go away?

BFAS is the most used document on the OCLC website, so we know it is really popular, people use it all the time. We have put a lot of work into the documentation. Many of you are aware that we’ve been revising BFAS for several years, incorporating RDA, adding lots of great examples, going through the entire document. BFAS documents the particular uses of MARC21 that are specific to the cooperative environment of WorldCat. MARC21 itself doesn’t take that into consideration, but Bibliographic Formats and Standards absolutely does.

So, no, we won’t be getting rid of it.

And if ELvl code 4 is obsolete should the WorldCat record be updated/enhanced?

We did change some instances of Encoding Level 4 to “blank” in the case of CONSER records, but we did not do the same thing for BIBCO records, the monograph records that carry the designation PCC. It’s something that probably needs to be discussed again. If, essentially, what was Encoding Level 4 meets most of, all the requirements perhaps, of current Encoding Level “blank” then maybe we should change them. We really don’t know at this point.

What field will be used to record the data that the record came in via batch process?

We don’t know yet. In his slides, Robert said that is one of the things we have yet to determine. When we make a decision and start implementing that, we certainly will announce it widely. It’ll be a future year when that happens.

Will the BFAS be extended to cover the numeric encoding levels?

The numeric and “blank” Encoding Levels are already described on the Encoding Level page in BFAS, and we recently revised that to change the text under Encoding Levels I and K, in particular, to explain that they will eventually go away, and that you ought to prefer use of Encoding Level “blank” and Encoding Level 7.

Are you giving any thought to running some kind of validation process over encoding levels other than M? Anecdotally the existing values are often wrong.

Yes, that certainly crossed our minds. If we develop criteria for assessing an M level record to decide whether it should be “blank,” or 7, or 3 it makes sense that we may want to do the same kind of thing on Encoding Level I. I’m sure all of us, at one point or another, have seen an I level record that was pretty deficient on detail. It makes sense to, perhaps, reassess some of those and end up with Encoding Level 7 or 3, rather than just mapping Encoding Level I to “blank”.

Have OCLC started chancing I levels records to blank levels

Not yet.

Is encoding level important to the duplicate detection and merge process?

Not in and of itself. What happens in DDR is that records are retrieved that look like candidates as duplicates, the data elements are compared in the records and then if it’s determined that the records do represent the same bibliographic resource they’re handed off to another process that we call Resolution. And what that process does is it takes a look at the coding in those records. It also considers the number of descriptive fields present and the number of holdings on a record, because that’s often an indicator of which record is better. So, you could have, for instance, an M level record that has 50 holdings on it and has 40 fields versus an I level record that has 5 holdings and half as
many fields. So between those two, even though on the surface it looks like Encoding Level I outranks Encoding Level M, where all the holdings ended up and the number of fields are more important a consideration in terms of retaining the record that appears to be most complete. So, in terms of hierarchy that we have, we give special consideration to PCC records, CONSER records, records from certain national libraries, but then for most of WorldCat they are as essentially viewed as being the same and uses this other criteria in terms of completeness: the number of fields and the number of holdings.

Can you explain why "M" is considered so important? Yes, it tells me that the record came in as batch, but why is this important to me (or you), especially since as a cataloger I can manually change the "M" to blank or 7, thus "losing" that information even now?

This is buried in the history of OCLC, and when the OCLC Encoding Levels were first put together, and someone at that time considered it to be important. As we look at it now, when we do work with records in WorldCat and are resolving problems with records it's a huge clue to us as to why something may or may not have matched because if it came in via Batch it's using our Batch matching as opposed to a human being doing the matching. It may explain why there's a duplicate record.

And in the focus groups we got the same feedback several times that it was important to know that a record had arrived via Batchload process. it's not so much that M is considered important, it's the information that the record arrived a certain way at least initially. We've finally come around to realizing the way we have been doing it for several decades is really kind of a bad idea, so let's fix it.

And to address the last part of the question, about that information that something has been Batchloaded into WorldCat can be lost now by an upgrade to “blank” or 7 or any other Encoding Level, and that's true, that’s been true all along. But we're trying to, in defining whatever it is that we end up defining as the new place within a MARC record, to record the fact that something has been Batchloaded into WorldCat that that information will be retained, and will be saved from then on.

Is this why M has been considered "Min" even though that's not really what the code means?

Originally we had Encoding Level L which was Full level added through a batch process versus Encoding Level M that was Minimal level added through a batch process, and of course decisions were made as to how to code those on the basis of a file of records, not the individual records, their differences within a file. We arbitrarily took a file and said, these look pretty good they're going to be Encoding Level L and these other records they're Encoding Level M. Of course, that doesn't really work out well when you receive a library’s file and it’s a mix of complete records and less-than-complete records. So, we ended up getting rid of Encoding Level L, there weren’t all that many of them, in favor of Encoding Level M because M had been used repeatedly over the years for the vast majority of files.

So, would you prefer we not change M to blank when enhancing at this time?

If you're doing an upgrade, you should change it to “blank.” It would be useful to do that. The question may be asked in terms of, well we’re losing the fact that it was Batchloaded, but that is the case on millions of records anyway. It seems to me that once a record is upgraded the fact that it came in through Batch initially may not be as important. The way that we use Encoding Level M now is often to diagnose problems, how did this record come to be this way, and it got added to the database without any human intervention. Once a record is upgraded, that means the cataloger has looked at it and made specific changes. I don’t think that’s a problem. You should go ahead and
change Encoding Level M to “blank” if you are enhancing a record.

I see authority linked headings that don’t get updated. Why does that happen?

We do have an issue with some headings not getting updated. It is on a list for investigation and a fix.

Is there a webpage that explains criteria for dupe detection?

There isn’t a specific page in Bibliographic Standards or anywhere else that explains Duplicate Detection and Resolution (DDR). There are documents however, and presentations that give you some additional information about DDR. Probably the most detailed account of, what I guess you could call the criteria for DDR is When to Input a New Record in Bib Formats and Standards, because DDR is based on When to Input a New Record and When to Input a New Record is largely based on what we do in DDR. So that would give you the best idea of what the criteria for Duplicate Detection are.

When controlled headings don’t update (I’ve seen it too) is there any more we should do than just wait longer?

The only options are to manually recontrol them or report them to authfile@oclc.org.

Why do field errors that prevent exporting records into a local ILS get through the validation/production process? Thinking particularly of invalid $2 codes and duplicate fields

That happens because those records that have those kind of validation errors did come in through our Data Sync process and we have some loosener criteria. They still counted as validation errors when they came in through Data Sync but we have three levels of errors and the least egregious level of errors, Level 1, we do allow those to be added to WorldCat, and then, of course, somebody has to fix them manually. But that’s why if we said, “no validation errors could get into WorldCat”, the vast majority of records probably wouldn’t get added to WorldCat through Batchload.

In the case of the invalid subfield $2 codes this has come up several times this year and we are looking forward to making a change where those would no longer automatically get added or potentially transfer from an incoming record to an existing record in the database; because they are so problematic and often the solution for us is to simply get rid of the subfield $2, change that 6XX heading to a second indicator 4 and then often that means that the heading ends up getting deleted anyway. Then what was the whole point of adding it? So, we are trying to fix this problem because we realize it does affect copy cataloging in a significant way.

We think that fix will go in later this year, we don’t have a date for it yet. It’ll be announced through the Validation Release Notes when it is ready.

Can you submit corrections to QC without 'item in hand'?

Yes you can, depending on what the error or the correction is, though, and where it is, if it’s a descriptive element such as the title correction or publication information, paging, that kind of thing, we require that you submit proof of the item so that we can make the change appropriately. Because we, obviously, don’t have the items in hand so we’re not able to make those types of corrections without the item.

Sometimes we are able to verify information through open source information on the internet, so I wouldn’t say that if you don’t have the item don’t necessarily report it, because we may be able to figure it out through other means.
We are doing some authority work and I am finding many eBook records which do not have the established form of heading for the author entries. And many others that have the wrong form. Is there anything that can be done to fix these?

Yes, you can report those to us. You can report the access point that was established. You don’t necessarily have to report individual records, but you can tell us that there are records with this certain form on them and we will make the corrections to the form of the name.

August 2020: Unraveling the Mysteries of a Merge and DDR Improvements

Topic-specific questions

August 4, 2020

Does the recovery process include correcting the incorrect coding that caused the merge in the first place?

No, it does not. Usually if we can discover what caused the incorrect merge, then we make corrections to the record manually after the recovery process.

We try to learn something from every incorrectly merged set of records. Obviously, if the incorrect merge was caused by incorrect coding, that's one thing. But, if it's something else, that suggests something more systemic or something that we overlooked or not treated better, we try to learn something from that and go back and do our best to build into DDR ways to avoid making the same mistake in the future, if possible.

While recognizing you all are super busy, would it be possible to let a requester know when an incorrect merge is fixed? I made a request on June 29, the record was corrected on July 23, but I didn't think to check the status until today.

Yes, we do try to let you know when the records have been pulled apart so that you're able to make any corrections on your end, or add your holdings to the appropriate record once they've been pulled apart, etc. Unfortunately, that doesn't always happen, but we do try to get back with you and let you know.

As records are submitted (manual) to QC to merge, does QC staff keep notes to help improve DDR?

We watch the process as we merge manually, and we have a match tool where we can input record numbers tool to see what DDR would do with the set of duplicates. We provide that kind of feedback to the DDR team. It depends on what you notice when you're merging and the time that you have stop and take a look then follow through with the investigation. As was mentioned earlier, we are continuously working to improve DDR.

Do you ever call in an item to see exactly what is being cataloged or do you just review the records in question based on what is in both records?

We're not a holding library, so we do not have access to materials. Sometimes we are able to view item information, including full text, from various websites, but that's not always the case. Consequently, if you report an error such an incorrect title, wrong paging, or bad publishing information, we do ask that you provide scanned copies of the item as
proof. That way we're able to make those corrections based on what is on the actual item. But no, we do have to take what is in the records unless we are able to find enough information for the item on the internet.

**Would you merge an AACR2 record with an RDA record?**

Yes.

**Is the list of what constitutes a significant change (that would trigger the DDR) available online?**

I don't believe that that's ever been online anywhere. It doesn't really change because we've used the same comparison points in DDR for many years. So, it seems like it's something that we could consider adding to the documentation in the future.

If you have a record that's online and there's a change to it, if the change happens in a field that we never look at, in terms of a comparison, then it doesn't make sense to necessarily put it through DDR because it's already been through that process in the past. We concentrate on those kinds of changes that are made in fields that we look at. So, if the wording in the 245 is corrected or the coding in the 245 is corrected, we compare the title fields, of course, so that's something that would then go into the DDR stream for processing. And then, it would be looked at seven days later, but it would be in that processing stream. Place of publication, publisher, changes to date. Whether it's the date in the fixed field, or the date in 260 or 264 subfield $c, changes to the extent, changes to the size. All of those are the kinds of things that would trigger a record going back through DDR.

**If we know there are a category of records that have a high number of incorrect merges (CIA maps where the only difference is the presence of relief), is there a way to get them unmerged en mass?**

If you want to send a list of the OCNs that have been incorrectly merge to the bibchange email address, we would be happy to look into them. If they have not been merged too long ago, we'd be happy to have them recovered. A lot of times with these when the presence of relief is the only difference once they're unmerged then supplying an edition statement in brackets to both of the records will help prevent them from being merged in the future. There have also been improvements made to DDR so the numbering in quoted notes in these types of records is taken into account. Sometimes when the only difference between some of these maps is the presence of unique numbering, if those are added as quoted notes in 500 fields, that is taken into consideration.

We were in close consultation with the maps and cartographic materials community in making a whole bunch of improvements to map matching, including things like looking for various constructions of dates in notes, especially in quoted notes, looking for unique numbering in quoted notes, or in non-quoted notes, various other things as well. It was also in consultation with the cartographic community that we changed the date, that is to say, we automatically do not merge records for maps that were published previous to 1900.

You may be familiar with the cataloguing defensively series that we've been putting together for basically the past decade. There is a cataloguing defensively presentation specifically about maps and it gives all sorts of hints about how to make a record unique so that it will not merge to another record that is very similar, but distinct.

**What's the time limit on recovering merged records? That is, are merged records not recoverable after a certain amount of time has passed?**

That is April of 2012 is the limit. Journal History keeps a record of transactions that have happened since April of 2012 and so anything that happened prior to that we would not be able to view or to recover.
There are large numbers of UKM records which were apparently batch loaded three or four times with variations which likely resist DDR, e.g., run-on 245 $a, edition in 245 $a, 260 with place and publisher transposed, garbled 300 fields. Could DDR be customized to target and merge these so that the best record in the set becomes a better bet for getting all of the records merged into a record from a different source? Or am I making wrong assumptions about these? (Why? These records triple or quadruple the work of manual heading maintenance.)

Yes, we have been talking about these lately. DDR is not easily customized. It's designed to deal with all sorts of situations that exist in bibliographic records, in terms of correct coding to incorrect coding and various kinds of issues with the way the data is formulated. But when it comes to these particular records, and what you're really talking about are the ones that have an 040 field that will say UKMGB, is that they are a result of a retrospective conversion process by the British Library. And the data is mixed around in fields in a way that DDR cannot handle. So, one of the more typical problems is that you see the paging in 260 subfield $a, rather than in field 300. And then the indication that the book is an octavo is also in subfield $a in 300, rather than in subfield $c. Those kinds of issues really do get in the way of DDR and have to be dealt with in a whole different way. We had a similar issue with records from the Bavarian State Library in the past, what we did was use a macro to look at different pieces of information and basically sidestep DDR and do a sort of a quicker evaluation of whether the records were duplicates and merged the lower quality record out of existence. Something like that could happen here, but we've also been thinking about possibilities of getting replacements for these records that we would perhaps match on the number that's supplied by the British Library in 015 or 016. That kind of thing could maybe take care of the problem once the data is cleaned up then possibly these records could be processed by DDR and merged. In a lot of these cases, we go looking at these records, look at the messed up record, go looking in the database for the same resource, and find that there is a duplicate - it's just that DDR couldn't match it to that record. So, yes, we are aware of these records and trying to do something to take care of them, but they'll probably be around for a little while longer.

If the same type of resource (e.g. HeinOnline cataloging) which are all batch loaded from the same institution, and have the same error (e.g. 260/264 with $a indicating no place of publication, but $b has a location and the publisher), if several of these have been reported, does OCLC note that this may be a repeated error and look for more of these from the batch? Or does OCLC require new error reporting for every item found.

I would recommend calling that out when you report it, so that we're aware. We don't necessarily go out looking for more duplicate records when we see a particular cataloging issue. Something to note is a limit to how many records will match. That limit is twenty records. Which means, if there are more than twenty records with this problem, even if we tried to send them through DDR, even if they are seemingly identical, it's possible that the twenty record limit is getting in the way of those getting merged. So that's always a good thing for us to know as well as maybe there's something that we can do to take care of these duplicates and fix the records in a different manner.

Is DDR flow automatic? Can anyone suggest a merge and how?

Yes, the DDR flow is triggered by significant changes, or if it's a new record added to WorldCat, We do have a process here that we can use to feed records into the DDR flow, but that's not something that's available externally. But you could also call that out, if you're reporting something and you feel that those records are identical, so they should be merged, it's something that we could get into the flow. If anyone does spot duplicate records, they can report those to bibchange@oclc.org.

This may be an issue for holding library to do their record maintenance. Since DDR is very dynamic, it is possible that the holding library may have obsolete Connexion records. Is a recommendation from OCLC to the individual library to update their records with the latest Connexion records. I understand it could be a local
issue for the individual library, but I just wonder about the OCLC’s recommendation.

I think this is talking about when you're getting WorldCat updates through Collection Manager. And so, whenever we are merging records here, then you're getting an updated record through that feed for your library. It is totally up to you, whether you change that in your local catalog or not but certainly, if you want to keep things up to date, using that Collection Manager, WorldCat, updates feed is a good practice.

Even if your holding was on a particular record that got merged to another, the control number is still indexed even though it's not the main record in WorldCat. The control numbers of the merged records are retained in Field 019 for this very reason.

How do you handle records of the same item with different languages of cataloging?

These are actually not considered duplicates. They're what we call allowable duplicates and are not merged. Duplicates that use the same language of cataloging can be merged. But if you have an English language of cataloging record and a German language of cataloging record that are duplicates for the same resource, those are not considered to be duplicates.

We use WMS. Do we have to update our records when they affect Collection Manager? I have reported duplicate records for journals we own in Collection Manager when the records were merged, were our E holdings transferred to the merge record?

Yes, when records are merged, all of the holdings are then merged into the retained record. If it's an incorrect merge we send them to be recovered, then once the records are reinstated the records are entered back into WorldCat as separate records with their respective holdings intact prior to the merge.

Hopefully, this isn't too specific a question for this platform. If so, just pass on my question. I am looking for ways to protect records against being identified for DDR when the difference is not necessarily represented anywhere but the 300. Ex: same title in two braille languages such Unified English Braille vs American Braille English edition. Since I am creating both records myself and have the opportunity, is there any advice on how to protect these records?

If the resources themselves do not have edition statements to the effect that you've indicated here, you can legitimately under both AACR2 and RDA add a cataloger-supplied edition statement that will differentiate the two records. If it's not stated on the item, you could bracket unified English, Braille edition as a 250 on the appropriate record and bracket American Braille English edition on the other record. And the 250s will be compared against each other and DDR will not merge them again. This is an example of the kind of thing that's dealt with in the Cataloging Defensively series. So, you may want to take a look at that.

What do you need when we manually report duplicate records? I usually use Connexion Client >>Action >> Report error. I usually put the message "Please merge with #12345678" (The "report error" command automatically sends the record number that you are viewing to QC.) Is this an OK way to report duplicates?

Absolutely. We will take reports for duplicates anyway you want to send them even if it's just a plain email to the bibchange address stating what the record numbers are and that they may be duplicates. The method you've been using is great. The window that opens when you choose Report Error does send us an image of the record as it appeared when you filled in the report.
Regarding e'holdings, are the merged OCLC numbers changed in the Knowledge Base?

The process that pulls merged OCNs runs nightly.

Non-DDR question: For bib records that have a 776 pointing to another version of the resource, if we change the form of the main entry in the 1xx, are we also expected to change that heading in the 776 $a to make them match?

It's not a requirement, but it would be good if that change were also made in 776 field. Also, if your workflow permits, it is best to call up the record that is cited in the 776 and make the change to that record directly as well, but it's not required.

A real quick way to update the 776 is to use "Insert from cited record" under the Edit menu in Connexion. You could just pop the OCLC control number in that field and update it that way.

When a heading in 1XX in a name authority record is changed, how long should it take for controlled headings that are linked to it to be updated. I've seen cases that take weeks.

Normally it should not take weeks. If there is some kind of issue within the NACO nodes and LC in terms of getting records distributed, there could be some delay in processing. On our side if we receive an authority record that's been updated by the Library of Congress, and the heading is changed once we load the record it will normally take 48-72 hours to get the change made across the database where headings have been controlled to that authority record. That's not to say that it's a perfect system, sometimes there are various kinds of issues. It may be worthwhile, if you've noticed that we have loaded a record or that if you're working in NACO you've made a change to a record and that change has not been propagated across the database, let's say after a week or so that you might send us an email to say, is there a delay or something like that? So that we can investigate what's going on because it is unusual for it take more than 72 hours for the process to complete.

A library which shall remain nameless has been adding local 856 fields to EEBO microfilm records. Could I send one example and you can correct the rest?

Absolutely, we do that kind of work all the time. So please send it to us, we'll take care of it.

Could OCLC look for 856 fields that have URLs that start with "proxy" and are clearly local URLs, could these be batch removed?

If the URL is one of these proxy URLs where the real URL is embedded in a longer URL, we have some coding in a macro to transform those into what should be the real URL, which then is oftentimes a duplicate of a URL that's already in a record that causes them to be collapsed into a single field. Requests like this for us, would mean, perhaps doing a database scan and running our macro through a set of records to try to clean them up.

If you have a set of records that you wish us to look at, send them to bibchange@oclc.org and we'll take a look and see if we can do some sort of batch processing on them to fix them.

August 13, 2020

Do you want catalogers to report duplicate Dublin Core records? Are they being deduped by DDR? Sometimes I see duplicates. Some are character-by-character duplicates and others have the same data, but fields are in different order. When I reported some Dublin Core duplicates earlier this year, I was told no action is taken...
because OCLC policy is that Dublin Core records can co-exist with regular MARC records, but that is not what I meant.

DDR does not process the Dublin Core records that you see come through as part of the Digital Gateway. So, they are not being de-duped, and we don’t merge those manually either. Instead, at this point, we essentially warehouse those records in the database because libraries will go ahead and harvest data. They get added to the database and then later they can be taken out and reinserted again. So, they’re sort of a different category of record than the traditional bibliographic records that you find in WorldCat.

We do not manually merge them because of the re-harvesting. If we were to merge them then they could possibly just re-harvest and another record be added to WorldCat.

How about DDR records in other cataloging languages? I deal with a lot of those that have $b spa and they are clearly dups in the same language of cataloging.

Yes, all records get considered for DDR regardless of the language of cataloging, other than the exceptions listed. We merge all different languages of cataloging, the rules do not really change how records are considered for merging it’s all the same, independent of the language of cataloging.

However, records will not be merged across languages of cataloging, so a Spanish language of cataloging record will not be merged to an English language of cataloging record. But within the language of cataloging they will be merged.

Is the matching algorithm used for DDR the same as used in matching for Data Sync? I’m wondering if there are discrepancies between these two systems. For example, could I send a record via Data Sync, resulting in the creation of a new record in OCLC (i.e., no match); then, days later, this original record is merged to a pre-existing OCLC record as part of DDR?

The algorithms used for Data Sync are different, there aren’t many similarities, but they are different, they have different purposes. So, yes, it’s very possible for a record to be loaded via Data Sync and then a week later be merged by DDR.

When you say that newly added records are added to the queue in DDR within 7 days, does that mean that they are actually evaluated by DDR within that time, or does it take longer for that to happen and records to potentially get merged?

It may take longer; it really depends on how much is in the flow already for DDR to work on. So, if we happen to have a higher amount of records that were added within a certain timeframe that may slow it down a bit. But it’s generally within seven days.

How does DDR work in tandem or separately from the Member Merge project?

With the Member Merge Project, the participants are actually merging the records in real time. So, the merging is happening instantaneously. They’re comparing the records and then going through the process of actually getting them merged. DDR is the automated process. The two are completely separate and different.

How many bibs does DDR evaluate on average per day?

Just looking at our stats, for example, in July they were seven point two million records that went through the DDR
It's usually between five and eight million records each month going through the DDR queue. So, we do examine a lot. Based on what was said, that means if a new record is added and a new OCLC number is created, then merged seven days later it shows which OCLC number is kept.

The OCLC number that's kept is the one that belongs to the record that ends up being retained. Whichever record makes it through the criteria in terms of the record retention hierarchy, is the one that's kept, and that number remains in the 001, the records that end up being deleted are in the zero one nine field. Most likely if everything is otherwise the same, you have a member record versus another member record, that was just added through Data Sync, probably the existing record in the database is going to be the one that's kept because it's been there longer and it had more opportunity to pick up holdings and possibly have been enhanced, but it could go the other way around. It's not necessarily the case, I mean, if the incoming record is a far better record, more complete, and the existing database record doesn't have very many holdings and is really sort of skimpy in terms of the description, the number of fields, etc. then the new record could be kept, but the number that ends up being retained is based on which record is kept.

Is there a table for automatic/manual merge? What it is the dup merged? i.e. is there a table for what kind of records we keep?

There is a hierarchy of records that is used in automated merging to, for instance, keep a Conser record over an ordinary serial record. When it is member to member record, we look at the number of fields that are present and the number of holdings, and then decide on which record to keep on that basis.

On the other side of the aisle, am I correct in assuming that you are still detecting duplicate authorities and reporting those to LC?

Yes, we report any duplicate authority records that are reported to us that are from users that are not NACO participants. We report them to LC on behalf of the library because only LC staff can merge/delete authority records as the LC/NACO Authority File is the Library of Congress' local authority file.

There's also a report that OCLC generates and sends to LC monthly for duplicate name authority records that are exactly the same.

We've found records with dozens of OCLC numbers in the 019 field are there that many duplicates?

Yes, there are unfortunately. We do have cases where records that have just slight variances that would not get triggered or caught by our process and then, therefore, get merged at a later time either manually. Or we have cases where they are identical. They get added via Batch and they are identical, and they do end up getting picked up with DDR and merged at that time.

It should also be mentioned that in a single DDR transaction there's a limit of twenty records being merged into a retained record and if it goes above that, we set it aside and the merge does not happen.
We are allowed to modify records to suit our local needs by including data that should not be added to the WorldCat record, such as coding subject headings as LCSH or LCGFT when they are not valid headings in either. Then the record goes through data sync and that "bad" data ends up in the WorldCat record anyway. Why?

They're actually what we call field transfer rules, but subject headings may transfer to the WorldCat record if there's a scheme of subject heading on the incoming record that isn't present in the WorldCat record. So, if, for example, the incoming record had a Medical Subject Heading (MeSH) and the existing WorldCat record had only Library of Congress Subject Headings (LCSH) then that MeSH heading would transfer over to the WorldCat record. If the existing heading already had MeSH headings on it, then the MeSH on the incoming record would not transfer.

We have worked very hard to improve what transfers and what does not transfer. So hopefully for the newer records that are being added, or loaded, through Data Sync that's getting better and we aren't transferring as much as we used to.

User comment: An example of that: Piano music coded as lcgft, which it isn't.

... suggesting that people ought not to be coding those things, lcgft, in their local system either. Probably better, if your local system will permit that, to code those things as local, when they go into your local system, if that works in your display.

It doesn't seem like a lot of people will take Library of Congress subject headings that they then intend to use as form/genre terms and automatically add subfield $2 lcgft, when, in fact, they really should be consulting that particular terminology, to make sure that it's there. It is sort of a problem for us, because our software just looks at that subfield $2 code, and looks at the record that is being retained in the database, either in a merge or if it's a case of a record that's coming in through Data Sync where we might transfer the lcgft heading, and if the retained record doesn't already have any lcgft headings, then it will transfer. So, you can take a record that was actually okay and sort of mess it up by transferring a heading that is not Okay. But yes, careful coding is always needed.

We just found records that had incorrect editions statements. What should be in the 538 field in records for DVDs. Is that because edition statements are a field that's transferred?

Is that because edition statements are a field that is, I'm assuming, that the local system wanted that more visible?

We were once told that the 250 is important field for DDR and that catalogers could "supply" 250 fields to prevent merger, defensive cataloging. I can imagine this on DVDs

The edition statement, field 250, is not a field that will automatically transfer, but 538 field is one that is that will automatically transfer, if it's not already present in the retained record.

More conversation about the DVDs and the edition statements: the person who asked the questions states that she was referring to two fifty fields that had things like widescreen and full screen.

And then someone else added: We were once told that the 250 is important field for DDR and that catalogers could "supply" 250 fields to prevent merger, defensive cataloging. I can imagine this on DVDs

Yes, that is true. You can add a cataloger-supplied edition statement in Field 250 to help prevent records that are otherwise exact in their descriptions from being merged.
The Cataloging Defensively series can be very helpful in giving you hints about creating records or editing existing records in a way that they will be distinguishable by DDR from similar records to which they should not be merged. So, you may want to take a look at the whole series of Cataloging Defensively webinars that are available from the OCLC website.

Why is that reports of dups for merge takes so long?

Unfortunately, duplicates are a problem and we do realize that, and we do have a substantial backlog of duplicates. We are working as best we can to get through them, but they do take time to go through as we have to analyze each one and make sure that they are duplicates and merge them accordingly.

I manually merged two bibliographic records (as part of OCLC Member Merge) and an incorrect 650 field got saved in the retained record that "validation" would have flagged. I then realized that the merge process does not perform a "validation" check before merging. Am I correct about this?

Yes, you would have to do a manual validation on the records or record after you’ve merged it, there’s no validation built into the merge process.

But if you are a participant in the OCLC Member Merge Project, you should always check, as we do when we do manual mergers, always check to make sure that everything that's transferred is something that should've transferred. You can clean up the record afterward the merge, and we encourage you to do so.

What about libraries use 830 field for their local collection title and add a subfield 5, the code of their library? The system simply accepts it. But we should not use 830 for a local collection title, right?

No, they shouldn’t. Those can be reported to us if you’re not able to remove the local series and you can report those to us if you’re not sure you want to remove it and we will take care of it.

Is there an error that is a common trigger for DDR? Just wondering if there is a particular mistake catalogers make that we should pay more attention to.

One example of that is when catalogers enter a record for the online version and forget to code Form of item: o in the fixed field, or in the 008 field. That can trigger DDR because the lack of that code makes DDR think that both records are for print.

DDR can get confused by contradictions within a particular bibliographic record. A contradiction between a 260 or 264 subfield $c and the Fixed Field date, for instance, or a contradiction between the place of publication in a 260 or 264 subfield $a, and the country code in the Fixed Field, things like that. So those are particularly important to pay attention to: contradictions within the record.

We have a comment saying Bravo for Bibliographic Formats and Standards Chapter 4, which has the section on When to input a new record. It helps to determine whether there are duplicate records.

We’re glad that’s helpful. Chapter 4 in Bib Formats and Standards is written to reflect what DDR does, and DDR is programmed to reflect what Chapter 4 states. They really are supposed to be mirror images of each other. They should both be doing the same thing.
Does DDR treat records coded dcrm_ differently from other records, or are they put through the same system?

Actually, DDR tries not to deal with records for rare and archival materials. There are 25 different 040 subfield $e descriptive cataloging codes including DCRM. If DDR finds one of those, it will set that record aside and not deal with it at all. DDR won't merge those records. We leave the merging of duplicate rare material records to actual human catalogers.

Comment: Multiple editions (often gov docs) published in same year with different transmittal dates often get merged if a 250 isn't supplied to keep them from merging

Actually, we do look at field 500 to pick up on those kinds of date differences, particularly in the case of government documents, you might have a date like that that's in quotes. So, if you had something that said April 15, 2020, and something else that said, May 22, 2020, we should be alert to that kind of thing, and be able to differentiate on that basis. Although it probably is a good idea to have field 250 in that case.

I've had multiple dcrmb records merged incorrectly multiple times until a 250 was added...is this setting-aside a new practice or must that have been done manually?

That could've been done manually. We have to look at the records in Journal History to see how they were merged if it was a DDR process or if it was a manual merge. But DDR does not merge rare materials records.

What is your preferred way for us to report dups? Using the error report feature in Connexion? Or?

Any way you want to get them to us, BFAS Chapter 5, Reporting errors will show you the different ways that you can submit them. But if you just want to put the record numbers in an email message and shoot it to the bib change email address (bibchange@oclc.org), we'd be happy to take those too.

September 2020: Connexion macros

Topic-specific questions

September 8, 2020

What is the character you use for a subfield delimiter in Connexion Macros?

The symbol can be generated by keying in: Alt+225. The symbol would be the Unicode symbol: ß, which would look like the Greek beta or the German Eszett. From there you can copy and paste it in the rest of the macro as needed.

It looks like I can create a macro to strip field 020, subfield $z and fast headings from my batches, which currently I strip manually?

Yes, the macro language will allow you to do both of these tasks with a macro. One way that makes creating macros easier is to use existing macros and tweak them to fit your needs.

Can you tell us about the AddAuthority6xx macro? What it does and what it is useful for?

The macro provides menu choices to add a 6xx note field to an authority record. A lot of the notes used in authority
records in fields 6xx have certain prescribed wording. As the wording for these notes could be rather complex, the
macro was created to allow catalogers to choose the note that they needed without having to remember the required
wording. It’s similar to other macros that provide choices on which option to input in the record, however, like other
macros in the OCLC Macrobook, this macro is older and may need updating.

When using text strings to add certain fields, why do the diacritics show up on the character following the one I
put it on. For example, in the word Español, the tilde is showing up over the “o” and not the “n”.

When using text strings to add certain fields, you will need to use the Unicode syntax for the diacritic to make it
appear properly in the record.

When we use macros that display dialog boxes where a user must make choices, we often get into a situation
where Connexion hangs and the system displays and populates the dialog boxes EXTREMELY slowly which
kind of defeats the purpose of using macros. I have reported the problem to OCLC and seen it discussed on
OCLC-CAT, but other than restarting Connexion I am unaware of any other solution. Is there any other fix
available for this problem?

We are unaware of any other work around for this. We know that there is a memory leak of some kind when using
the macro language; for example, if you use the same macro over and over in the same session, but the only current
workaround is restarting Connexion.

Can I use OML (OCLC Macro Language) to insert today’s date in a text string? I’m specifically wondering if I
could use it to insert “OCLC, [today’s date]” in a 670 when I’m recording usage and headings as found in OCLC
for NACO work or is that already in AddAuthority6xx?

Yes, that feature is already available in the AddAuthority6xx. It will also insert the current date when citing a
database.

It seems like the current text of “other unmediated carrier” in the run33x macro for an unmediated carrier type
338 does not validate. Are there plans to update the drop-down text to “other” which does validate?

We will look into fixing this.

Is there any place people are pooling their own macros? I know about the Walt and Nick links, but how can we
share our one-offs?

OCLC-CAT and Get Hub are great for sharing Macros.

Can you take OCLC WorldShare questions?

If we cannot answer, we can direct you to the correct source for help.

Is there a way to resize the macro screen list to show all the macros you have available?

No. While you do have a slider bar, you cannot resize the window.

When printing labels (regular printing, i.e. no macros), the diacritics come out in the wrong place. How do I fix
that before printing? We delete the diacritics before printing the labels, then write them in on the printed labels.

This might be a good topic to put to OCLC-CAT for help.

We delete the diacritics before printing the labels, but this is certainly manual, time-consuming labor. An
alternative is removing the diacritics in the record before printing. Maybe a macro could remove the diacritics
before printing?

Yes, a macro could be written to delete all diacritics from the bibliographic record before printing the labels, although
you would want to do this as your last step.

Is there a way to bind macros to a keystroke command, so that you don’t have to go into the macro list to run
them?

Yes, you can assign using Keymaps or UserTools from the Tool Menu. Information on setting up Keymaps and
UserTools can be found in the OCLC Connexion Client Guides Basics: Set Options and Customize.

I sometimes have problems copying macros from a shared network drive to someone’s personal drive (or PC). I
get error messages about the syntax of the name or Macrobook or that the macro already exists, etc. It's
particularly difficult when I try to copy a macro into an existing Macrobook. Is there any online documentation
for this?

No, there isn’t any official online documentation that we are aware of. However, within Metadata Quality, we share
macros by either sharing the whole Macrobook or copying and pasting the macro text into a plain text file and sharing
the text file. Some macros though may be too big to copy all of the text. If you have a macro that is too big to save
the text in a text file, Walter F. Nickeson created a macro, MacroBookInspector, that will allow you to copy and save
the entire text of the macro to a text file for sharing.

Some macros are available in Record Manager. Are there plans to incorporate more macros to Record
Manager?

Macro functionality of some macros that we’d previously done in Connexion Client has been built into Record
Manager. Currently there are ongoing investigations into what other functionality can be incorporated into Record
Manager as well.

How long did it take you to feel confident in writing macros?

(Robin) While I feel confident writing macros for what I need to do, it took about a year to learn and I still refer to
reference materials and am constantly learning new things. The macros take time to create but they saved time in
the long run.

(Robert) Macro writing is a constant learning process. It’s based on Visual Basic, so you can often find general
examples for what you are looking for online. For example, sorting a list. A good way to learn to write macros is to
take another macro and clone it. Then play around and modify to accomplish what you need. Using an already
created macro is also a good way to save time.

It would be great to be able to search/display all records that have had actions taken on them by a certain OCLC
authorization. For example, to see all the records that the cataloger has replaced during a specified time period.
I know we can see the number of records using OCLC statistics, but I don't think there is a way to get a list of
the actual records.

No, there is currently no way to see the list of the actual records based on an individual authorization.
If all you want to do is add a variable field to a record, the Constant data function works well.

Yes, constant data is great to use for the same edit to all records in given list of records.

September 17, 2020

What does "have the fixed field turned off" mean?

In the Connexion Client, the Fixed Fields can be displayed with the textual field name next to each code(s). For example, “Type”, “BLvl”, “Dtst”, etc. You can also have it turned off where it’s a single 008 field with all of the codes listed in a flat variable field. There are no labels when using this option. You can change this by going to the View menu, clicking on OCLC Fixed Field, and then choosing one of the options.

Is there a place where we can see other's macros?

There are a couple example repositories in the presentation. You could also exchange macros and ideas using OCLC-Cat.

Does anyone know where to get the NikAdds macros?

Joel Hahn wrote that macro. You can find it at http://www.hahnlibrary.net/libraries/oml/connex.html.

I often use the Cyrillic to Roman macro. Are there other macros that OCLC is aware of for transliteration?

We are only aware of Joel Hahn's transliteration macros, which covers many different scripts including Greek, Hebrew, Korean, etc.

What's the difference between Macros and Text Strings?

A macro is the entirety of the script that you want to run on a record or on multiple records. A text string was the text inside of the quotes in the example shown in the presentation. A text string can be read into a record by a macro but the macro is the whole thing together.

Can macros be nested?

You can have a macro that will call another macro. You can also embed what you need from one macro into another macro instead of having to constantly call the second macro.

NikAdds appears to be a good example of an answer to the question of calling another macro from within a macro.

When I use the generate authority macro Connexion Client keeps freezing. The add usage and add 3xx templates are "stuck". Any ideas? I usually have to force Connexion Client to close and then re-start it.

This isn't a problem with the macros themselves but rather with the macro language and the Connexion Client, which can leak memory. This can freeze up your Connexion Client session. Presumably, more memory in a computer would help, but there is currently no other solution, than to shut down the Connexion Client and restart it. This slowing down and freezing up of the system usually happens when running a macro repeatedly on the same session.

Comments from attendees:
I also suffer from the computer freezing problem. This started when my computer was upgraded to Windows 10 (it is an older computer not designed for Windows 10). It does not happen on PCs built for Windows 10. I have to avoid all macros that initiate menus.

The auth generation macro can appear to freeze when a menu window gets stuck behind another window. Nothing happens until the menu is dealt with.

The freezing of the screen seems to be a problem when you have Connexion open Fullscreen. If you decrease the size of the window, the problem goes away most of the time.

On Freezing and Window size: Not for me. In my experience, it does not matter how big Connexion is sized on the screen. It is possible to find a second instance of the running application in the taskbar, select it, and force it to stop. It is a little tricky sometimes, but it can "stop" the macro window, so you don't have to restart Connexion or to shut down and re-boot.

Is OCLC still creating and revising macros for Connexion for member use?

Yes, one new one is the punctuation macro and a recently revised macro is the macro that generates field 043 because the names of countries have been changed overtime. We could add more if they would have widespread use.

The OCLC Guide ("Basics: Use Macros") has a lot of info on the CS functions, but is there a document that talks about the Visual BASIC (VBA) component of OML? It doesn't seem to work exactly like VBA.

You would have to go to documentation outside of the Basics: Use Macros document that covered VBA. While most VBA commands do work, not everything will work exactly the same. There is also the situation where some VBA commands are newer than the set of commands available when OML was created. You can often find VBA documentation and commands by searching online and these examples can be very helpful when creating your own macros, especially when the command is more complex, such as sorting a list.

Walter Nickeson added that OML seems to be virtually identical to IBM's CognosScript.

When would you use constant data to insert a field or several fields, rather than a macro?

Depends on what fits best into the project your working on and your workflow. Sometimes inserting a text field works, sometimes constant data works, and sometimes macros work.

I really like the 33X macro but find that it still leaves blank 33X fields behind that have to be deleted. Am I missing something? Is there a way to auto-delete those once the macro runs?

This might be a prompt from the Workform. If you create a record using a Workform, then the prompt for the 33x fields will appear on the screen. Running the Add33x macro will not remove the ones already in the record from the Workform but will add the appropriate new once into the record. All of the prompt fields in the Workform will be removed when you do a Reformat command.

An attendees list of favorite macros, which were written by Walt Nickeson and Joel Hahn:

- CopyDateTo264_4201601
- Add040$beng
When I've used dialog boxes, I can't seem allow the user to choose from a menu or enter a custom entry. Is there a way to do that? To clarify, the box will accept the user's choice from the menu options, but it won't accept an input that isn't an option.

Yes. Dialog boxes are one of the more complex things to do with using Connexion macros. You have to define the box, define all of the variables that are going to capture the data from the box, then command to display the box as a specific point in the process. Because of this, when a dialog box isn't working properly it could be that either the macro was corrupted or there could be something missing. Walter Nickeson offered his help with dialog box creation if anyone needs assistance.

I tested using the Latin2Cyrillic (Russian) to a record in my save file. The Russian script looks like it had invalid characters in it: А страы дроп оф блоод / ´с Росеанна М. &#x0220E;&#x0220E;ите. This is not a record I want in Cyrillic; I'm just playing with it now to see how it looks, the ampersands look wrong this is what I was transliterating: A stray drop of blood / ´с Roseanna M. White.

Maybe something that you can fix so you may need to fiddle with macro. The sequences should be an ampersand, pound, and "x" followed by four characters and end with a semicolon.

Attendees also added:

- The problem with the Russian record may be because the transliteration is incorrect.
- The transliteration macros don't work for languages they are not meant for, like English.
When I'm authorizing bib headings, I can do everything with keystrokes except get to the next record. That seems to require a mouse. Is there another way?

So, if you are controlling headings in bibliographic records and processing a set of bibliographic records in a save file, then CS.GetNextRecord ought to work. You can also use Keymaps go forward or back in list. To do this, go to the Tools menu, click on Keymaps, check the “Menu item” option and select ViewNavigateRecordsandListsForward and ViewNavigateRecordsandListsBack to assign Keymaps.

Will a macro language be planned for Record Manager?

Adding OML to Record manager is not possible, which is why some of the most popular macros have been built into Record Manager. We do continue to add to Record Manager functionality so if you have an idea, you can submit an enhancement request via Community Center.

Can we use ISO 639-3 language codes in the 008, or only MARC language codes?

Only MARC language codes in field 008 (or the Language Fixed Field) but you can use ISO 639-3 language codes in field 041, with the second indicator coded 7 and the appropriate code in subfield $2. OCLC is also trying to make more use of field 041 in our discovery system so that we have those more granular representations of languages that you can put into field 041 but cannot put into field 008.

Attendee comment:

Just wanted to shout out this webinar from Georgia Library Association that went in depth on macros, text strings, constant data, etc. I hadn't had time to delve in before recently and it was a great introduction: https://vimeo.com/440363659.

October 2020: Linked data, and the road to learning about it

Topic-specific questions

October 6, 2020

Please define "serialization."

Serialization is basically a way of formatting the data. It’s kind of the way that you code it in the background.

Attendee comment:

Simply put RDF data can be output in different formats, like JSON, JSON-LD, turtle TTL, etc. for Web services to ingest and process.

Machines negotiate content for delivery in HTML so humans can digest. It depends on the receiving service requirements; structural data can be queried, and output based on its requirement for processing to meeting the need.

JSON, JSON-LD and TTL are supposedly more human legibility friendly. Alternatively, some browsers may have
For non-coder catalogers, this will be behind the scenes and we'll be able to interact with user-friendly interfaces—correct? I hope so

Catalogers will be using a user-friendly interface as opposed to looking at all that coding.

It would be something like when you look at Wikidata entries. I highly recommend joining in the LD4 Wikidata affinity group, it's a good way to get an idea of what an interface might look like, that a cataloger would work in. But you wouldn't necessarily be working on the serialization side of things. In a sense it's like the MARC format, in that Turtle code, RDF, RDF/XML, etc. is just a way of coding the information so that the computer can do its work in the background. It wouldn't necessarily be what is displayed for the end user, or even necessarily what is displayed for the cataloger within the linked data.

Can the presenter talk about the (current or future) relationship between MARC cataloging and linked data? i.e. how you see the relationship, etc.

What we've been doing is prepping that MARC data to be used within linked data and as we transition more towards the linked data world, we eventually could leave MARC behind completely. But it will take some time as we transition and again, the subfield $1$, subfield $0$, those will help us translate that MARC data into linked data much easier.

Then focusing on the authority records some of the things we're learning is all of the information that's used to create authority records is in no way comprehensive about the “thing” being described. So, what does comprehensive really look like in terms of linked data? If you're describing a person, how much about that person do you want to know? And we're looking at it really from two ways: 1) what do you know about a person, e.g. their birth date, their death date; where they may have worked, etc. But 2) we're also looking at it from the bibliographic side of: Do we know everything they wrote? Do we know everything that was written about them? So, in terms of authority-ness , we're really limited in some ways, only to those people who wrote things. We're very light in the authority file on subjects for persons. We have the prominent people, but if you really look at… For instance, one thing I enjoy listening to is CBS Sunday morning. They always have this fascinating segment that's called A life well-lived. And I got really curious one day and started trying to find out, for all these people who had lives well-lived, their fascinating history. None of them were represented in the authority file. Well, that seems really odd, and yet many of them were in wiki-data. So, there were works that they contributed to, in terms of their life, but we're not able to capture that in authority records. There's a lot of room for growth, in terms of what we can do to find relationships between people who were not authors and works that represent their life well-lived.

The major block to adopting LD (linked data) in libraries is the huge amount of legacy (unlinked) data we already have. Are there strategies being developed to address this issue, especially for libraries with small staffs that couldn't possibly update their existing data manually?

A lot of work in linked data so far is in the academic library world. How do you see how public libraries will benefit from linked data?

There are clearly some people in the chat who are advanced at linked data policies, and procedures, and just how it all works together. One of our concerns with presenting this topic, though, was how do we reach the person who’ve been aware of the linked data stuff, but not necessarily paying too close attention because there hasn’t been an effect on their day to day work. And OCLC, and several of the other vendors, there is work being done on creating
the interfaces for you to use. So, you don't need to be a programmer or know RDF or that kind thing. And you'll make
use of linked data because the underlying structure will change and create all of these or allow you to transverse all
of these relationships. The LD4 community and the LD4 wiki affinity groups are looking into these interfaces. OCLC
is looking into the user interface in terms of how a cataloger would actually use all of this data. The stuff to do right
now, if you are in a place to do that is to, for example work on cleaning up the subfield $0 the subfield $1 in your
MARC data, if it's at all possible. If it's not possible, we're still open and sharing our data. Yes, we have a
subscription, so I'm not going to downplay that at all, but we're still based on the fundamental cooperative cataloging
model so no one will be left behind, so to speak.

Are the major ILs vendors, especially those that serve the public library market, involved in this work?

I believe that Innovative Sierra is working on Linked Data. Xlibris is also working on incorporating linked data into
Alma and Primo.

What is the level of adoption of Linked Data in libraries? How are they using LD (linked data) to meet their
missions?

I know that we've worked with several groups that are doing that. There are several libraries involved with the LD4
groups, that are working with the different standards, things like that. Right now, a lot of it is just experimentation and
developing the infrastructure to be able to use the data.

How do you respond to those in the cataloging community who are skeptical about RDA, BIBFRAME, and the
actual implementation of linked data?

What you can say about RDA is that it's an evolution in our cataloging instructions that is better designed for
transitioning to linked data in the future. Because one of the changes was an emphasis on actually coding
relationships. Under AACR2 we didn't supply, in terms of MARC, subfield $e, with relationships between authors as
access points in the same way that we do now under RDA. So, with all of that data specifically coded in our current
environment, it's the kind of thing that we'll be able to map forward in the future into a linked data context so that it
can operate on the web.

With BIBFRAME and implementation of linked data, some of this is still remains to be seen how it plays out. Some
healthy degree of skepticism is good, because it will get those questions answered that need to be answered in
terms of How does this affect me? How can I help my systems? What's the benefit to me? etc. So, having questions
like is always good.

I think the skepticism comes from the fact that this transition seems to be taking a lot longer than I thought it
would. Any ideas on why that is?

I think it's taking longer because it ended up being harder than we originally thought it might be because this has
been in the works for 10 years plus probably. We've made a ton of progress in the last 5 to 10 years and I think there
are going to be breakthroughs with actual practical use. OCLC, the LD4 community, the other vendors that are
working with us in the next year or so as, as we really home in and move this all forward.

Or is it taking longer because discovery services we use are lagging behind?

That's interesting, though I don't know how well we can answer that since most of us are catalogers and our primary
focus isn't the discovery system. There's definitely that piece of it: how the discovery system works for the end user,
not the librarian, not the cataloger, but the students or the public that come into our institutions. How are they ultimately going to use this and make those connections? Those are questions that still need to be answered, in the grand scheme of things.

**Participant response: I think it has to do with legacy data.**

There is definitely something there. The goal with BIBFRAME was not to leave MARC completely behind and start fresh. There are aspects being ported over because that's how it is in MARC. Once we get into more of a linked data environment, it'll be good.

I also wonder if a part of the challenge isn't just as a community. We are limited by budget. We are limited by training, et cetera. Moving to a completely new dynamic infrastructure is a big shift. Not only do we have to understand it, but we've got to persuade budgetary constraints that this is a good use of their funds, knowing that we just don't have unlimited funding anywhere.

**Participant comment: It is hard to "sell" the idea without being able to demonstrate it effectively in a discovery/user facing system.**

Absolutely agree. And I think that's where some of the smaller libraries who aren't involved in the establishing rules, standards, etc. are struggling because they haven't been able to see the progress etc.

**Participant comments:**

Good point in that the reduction of cataloging staff during the past 5-10 years has been significant in all libraries and sectors.

My library participates in the Library.Linked project with Zepheira, which is a good first step for public libraries to get into linked data.

Wikipedia and Europeana effectively uses LD (linked data), at least that is my understanding.

Those are good communities to check out.

How useful can LOD be for increasing and diffusing knowledge? Having structure data readily available for query services, e.g. SPARKL, when there is a need to generate Web pages for a trending topic, e.g. CoVID-19, the process is much faster. Output data is constantly updated as the queries are being conduct when a user clicks on the link. [https://sites.google.com/view/covid19-dashboard](https://sites.google.com/view/covid19-dashboard) PS: The site was up in April 2020

Someone in chat points out that having the structured linked data, the linked open data, is increasingly good for diffusing knowledge. Being able to have these query services and the SPARKL end points, especially in our current environment, where most of us are working from home and lockdown, away from our normal infrastructure this sort of processing is so much faster than having MARC data locked up in our different MARC repositories, our different MARC silos.

She also points out that the output data is constantly being updated as the queries are being conducted so you don't have to worry necessarily as much about as stale data.
Our bibliographic and authority data is incredibly nuanced, especially in communities like the Rare Books community. Being able to map data between MARC and LD, and back again, without data loss is proving very complex.

Going back to why this is so incredibly complex, someone mentions being able to map MARC and linked data, especially outside of the normal monograph book cataloging, and specifically with the rare book community being able to map data between these and back again, without loss is proving to be very complex.

This is true of archival material as well since the MARC record doesn't deal well with collection-level information. Making sure that that information doesn't get lost and the contextual notions, ideas that are available within the description of those records makes it challenging to put a linked data wrapper on it.

How then will legacy data be "updated" or move to linked data environment? or will we have 2 systems running parallel to each other? AND BTW, a lot of people still not clear about linked data.

Absolutely, people are not necessarily clear about, or sold on, linked data, but I think as OCLC and other groups, like LD4 and PCC (Program for Cooperative Cataloging) as they continue their investigations for this that it will all become clearer.

As far as how legacy data will be updated or moved to the linked data environment: it has been something we've talked about a lot. There are certainly challenges. Knowing that when you look at a field in a MARC record, it's comprised of different subfields when you express those different subfields in data, you're looking at different properties. You've got to be able to pull apart the pieces and then be able to dynamically update them going both ways. It's certainly something we've been looking at. We understand some of the challenges. We have not solved the entire puzzle yet, but we certainly understand the need that those two expressions, if you will, need to have a relationship and how to maintain that is definitely going to be a challenge.

Of course, well-coded MARC data is something that will transition to linked data much better than the case of MARC records that are poorly coded and incomplete.

One reason we're having a hard time moving forward is that many are still trying to understand linked data through a MARC lens. Thinking out of the box and getting away from a "record" concept might help. One of my favorite linked data sites is here.

There seem to be a lot of systems/interfaces for linked data now and lots of experimenting is happening now. Is it likely that just one will eventually surface and we'll all use it, like MARC?

In terms of linked data, aren't all the instances where an authority has been linked to a bib resource also part of the LD (linked data) information about that entity?

We do look at this through the MARC lens and part of that is we don't want to necessarily lose all of this very rich data that we have trapped in MARC. But at some point, that does inhibit us to some degree. The linked jazz network is good for doing some exploring, to go through and see how Jazz is all linked together with all the different people. It's a very nice interactive website.

This is one thing that helped me get out of the MARC lens so dramatically was when I was asked to look at MARC data, because I was very focused on understanding these are the elements in a MARC record. What would they look like in linked data? What are the properties? etc. And when I got done someone I was working with at the time, looked at me and said: What question are you trying to answer when you look at a MARC record? What are the
questions you're asking yourself when you look at it? Are you asking: What is the title? Who was the author? And that sort of helped me break the MARC-ness bias I have, no matter what the approach, still, what question am I trying to answer? If we can start thinking about linked data as not necessarily how does it equate to a MARC bibliographic record or a MARC authority record, but what do we need to know to answer the question that were are posing? What are you trying to do? Perhaps, if we start looking at it like that, it might help us break out of that detail level of subfield $a, subfield $b, subfield $c. It gives us a better way of looking at what are we trying to do to help our users.

Participant comment: LC is creating BIBFRAME (which is LD [linked data] centered), and they have been able to convert MARC to BIBFRAME, and they are trying to convert MARC to BIBFRAME.

If you're interested, it is a good site to explore. They actually provide a good visualization of how BIBFRAME would look. How a MARC record would look in BIBFRAME. And then going from the BIBFRAME record to MARC. So, it is a really good tool to check out and play around with.

Attendee comment: I believe that LC is working on their backwards conversion so that we can work in two systems, per se

Where is that PCC document about $0 $1 in authority records?

The FAQ for the URIs in the presentation notes breaks it down and it does a pretty good job of explaining the difference between the two and the purpose of each one. Bibliographic Formats and Standards has information about the control subfields, including examples. The MARC documentation also does.

I find records with wrong 520 fields and youth headings attached to non-youth materials. I've been deleting them and updating the WorldCat record, but should I report them instead?

We would appreciate you reporting them, that way we can see if there's a bigger problem, and then find other records that may be involved with that.

Do we have any sense as to why this is happening right now?

They could be coming from merges when fields transfer, or they could also be coming in via Ingest.

If we report records with invalid 007 values that block replacement, would reporting them prompt a batch-based correction? when appropriate

Yes, we'd like to know about those kinds of situations where there's a problem that's widespread, because that is the kind of thing that that would lend itself to some automated fix. Some are easier than others, but certainly cases where you have messed up 007 fields that are getting in the way of being able to do replaces on records, that's something that we would like to take care of across the board. So, yes, please do report that kind of thing.

October 15, 2020

Will OCLC be offering a cataloguing interface for cataloguing in BIBFRAME? Something like Record Manager?

What we're working on is the SEMI project that was described in the presentation. We're working on the interface for that. Its relationship to BIBFRAME is something we can take to others. The relationship of SEMI to Record Manager
is certainly something we're thinking about, but right now we are trying to keep them separate in our approach to ensure that we address the needs and the user stories associated with linked data. And then we can look back at Record Manager to determine similarities, differences, and that type of thing.

**OCLC may or may not implement BIBFRAME?**

Everyone is talking about BIBFRAME as we're looking at SEMI, so, in no way are we saying we will, or will not implement all the discussions related to BIBFRAME. We are using all of the knowledge in the community from BIBFRAME and other sources and discussions to guide us in our thinking and our understanding. And there's a lot of information on the community site as we're working with the User group for SEMI to ensure that we are engaging the community and understanding their needs. And a lot of them are also involved with SEMI and other projects. So, there is a close relationship between what we are trying to build, and the standards that are under discussion in the community.

**Attendee comment:** ALA Fundamentals of Metadata had a good introduction to metadata and some discussion of linked data.

**Will the result of linked data be open access knowledge graphs?**

That too is under discussion. There are a lot of conversations going on regarding accessibility and what it means to have linked open data for WorldCat, understanding subscription models, and that sort of thing. So those are ongoing discussions. And again, I think a lot of that information will be made available because that too is a topic that has been put forward to the advisory groups on the SEMI team, and its users to get input on how people are thinking about who should be able to see what, what should be linked open data, what should be more guarded for OCLC members. So, a lot of discussions about that are in play.

SEMI, to reiterate, is the Shared Entity Management Infrastructure project that is funded by the Mellon grant, and it is underway now. The end of that grant will be at the end of December of next year, 2021. So, we expect by the end of that grant to have an interface for entities.

So, it won't be an interface for a bibliographic record, it'll be interface for just pieces and parts that are entities. One thing we're looking at is the different entities. We are creating what we're calling a minimum viable entity description. We're using the properties and classes to guide us in determining, not unlike some of the forethought that was going into Bib Formats and Standards, to determine what fields are required, that sort of thing. We're taking a very similar holistic approach to understand what properties we think are needed to be able to describe a particular type of entity and then growing the interface around those rules and thinking.

**Any comments on ORCID? I have worked on that in a remote project.**

Certainly, we're aware of ORCID a lot of research folks have been involved with that project. And it is one of the identifiers that we're looking at incorporating as a property.

**Can you address who would have editorial control over linked data entities?**

And again, I think that goes back into the interface that we're building, looking at that subscription type approach to modeling of who would have access. We're not trying to build the bibliographic infrastructure, and I really want to make sure that people understand, in no way are we looking to rebuild that. We're looking to that as guidance as we
make decisions. All of that is still very much under discussion. One of the things that I think we all know from Wikidata is that there are a lot of people in the world who know a lot of things about particular types of the information who could easily add statements and claims that they just know because of their education and their familiarity with the specific topic. We're trying to ensure that we provide a way to ensure that everyone can contribute to OCLC's linked data in the same way that we've seen the community build the Wikidata. So again, those conversations are very much under discussion, but we are definitely looking at ensuring that people can contribute claims and statements as they are aware of that knowledge and making sure that we keep that open so we could share that information. That's the whole point of linked data, to share what you know.

If you're using Wikidata now and you are adding statements to Wikidata, I think adding to our entity data, once it's ready or once it's ready for people to edit, will be very similar. We are definitely using Wiki-based infrastructure as our underlying technology, so many of the same concepts and some of the look and feel is very much like Wikidata. But we are trying to ensure that we fit it to meet library needs and the community.

**Any plans for WebJunction courses on LLD [library linked data]?**

Please see their website: [https://learn.webjunction.org/](https://learn.webjunction.org/)

**My understanding is OCLC has "published" linked data, what does that actually mean, and does OCLC know how anyone has used that published linked data?**

As far as I am aware, publishing linked data means making it publicly available.

One of the ways is the FAST Linked Data Service. [WorldCat.org](http://WorldCat.org) has some published linked data. VIAF® (Virtual International Authority File) is considered published linked data as well.

We know lots of people are using FAST, there's evidence that people are using the linked data OCLC has published and linking to it.

**Why is OCLC not going to BIBFRAME till now to get more benefits from Linked Data?**

A lot of it, as noted in the presentation, is that OCLC has spent a lot of time working with linked data, and again, we are certainly not ignoring BIBFRAME, but we're also drawing on our own research as we explore the work with identities, with CONTENTdm. There's a lot of knowledge there that we are using to help guide us, and all the user communities’ feedback that we have from those projects. So, again, we are in no way ignoring BIBFRAME. We're just trying to include everything that we have learned as we look at the other standards and discussions going on in the community.

One reason is that BIBFRAME doesn't actually handle RDA well. It is less full than RDA in RDF. It doesn't follow the LRM model.

There's still a lot of development going on with BIBFRAME and the Library of Congress is still experimenting with it as are many other people. OCLC has pledged to have a way to, in the future (no dates associated with this) to ingest BIBFRAME data. We will talk about that widely once we're at the point of figuring out what we're going to do with that.
I believe there are more than one format (style) for linked data. What is the best one to start with? RDF/XML

They all bring their own set of pluses and minuses. The resources listed in the slides help to explain it better. It really depends on how comfortable you are with coding and what style you like best: Turtle, JSON, RDF/XML, N-Triples, etc.

I continually get turned down for courses that have the words “fundamentals” or “introduction” and “metadata”, so I was wondering if it would be helpful to someone who has been cataloging for 30 years, but in MARC format, not BIBFRAME or RDF. My library administrators consider me an expert in metadata, so I don't get to take these kinds of courses.

There is a great linked data course on Coursera, Web of Data - quite advanced, I enjoyed it

How can OCLC help subscribed libraries to transform their data to linked data in the WorldShare Management Services (WMS) ?

We’re not familiar with any discussions, but it certainly doesn't mean there aren't any. We’ll just assure everybody that OCLC intends to support MARC for years to come. We don't have an end date on that. And we think the evolution to linked data will be just that, an evolution, rather than a revolution. So, there won't be a hot cutover at any point. We'll keep you posted as we have new developments.

November 2020: OMG, there are so many 5xx fields, what do I do?

Topic-specific questions
November 5, 2020

Should 501 be used for ‘reversed print’ books, containing 2 distinct titles? And if the reversed part is just a translation of the same work?

Under AACR2 and the LC rule interpretations plus OCLC’s policy with regard to cataloging resources as issued, if you had more than one title issued under one cover, you would create one record to represent the whole thing. If you had a back-to-back situation, which is typically a translation, you would create one record for that, transcribe both titles in the 245 field, and make an additional title in 246. Field 501 is mainly reserved for rare books. Under earlier rules, i.e. AACR1, field 501 had also been used for sound recordings when you had one work by more than one composer. You would create a record for each work and link them with the 501 field. That’s still allowed under AACR2 and RDA but is not the standard practice. Under current practices you would generally create a single record with the multiple titles in field 245 with subsequent works in 7xx fields.

Does field 502 transfer when merging?

Field 502 does not transfer.

For the discography examples in the 504 field, shouldn’t the pages be enclosed in parentheses like a bibliographical note?

The typical bibliography note should normally follow the formula “Includes bibliographical references” followed by parens with the page numbering. If it’s some other kind of note, you are using the word “Discography” as a caption, so you would not include parentheses in that case.

Does field 504 transfer?

Yes, field 504 does transfer when merging.

Does QC have any advice on alternatives when note fields don’t display in Discovery, for example language notes in field 546? We’re a WorldShare library, and we’re concerned that not all public notes are displaying for users. Our current approach is to continue as per BFAS and work on enhancement requests for Discovery. But are there different fields or different approaches we could look at? We’ve already considered LBD but these are fields that are already valid in the WorldCat record and LBD only displays in our local instance.

Our colleagues that work in Discovery are in the process of re-evaluating what fields should display and how libraries might be able to customize that. Jay and I are consulting with them on that and those conversations have just started. Submitting enhancement requests to the WorldCat Discovery Community Center is definitely recommended.

Does ‘No Transfer’ mean only when it is on the non-preferred record but remains if it is on the preferred record?

If we are merging two records together, and this applies to both the automated and manual process, we have the preferred record and the record that is merged into it. If one of these fields is already on the preferred record, that is the field that will be kept. However, if the preferred record does not have the field, for example field 504, and the record being merged into it does, then field 504 will transfer to the record being retained.

Will BF&S be updated to include the transfer information when records are merged?

We will be discussing including field transfer information into BFAS in the future.
Does the closed-captioning note now go in field 532?

Yes, it now goes into field 532. We will update the example in 546, which was the previous practice. There were two accessibility fields that were added to MARC 21 fairly recently, field 532 which is the accessibility note and field 341 which is the accessibility content. So far there isn’t very much official guidance on using field 341. There are no standardized vocabularies that have been established yet for the 341 field. The 532 field is more free text and although has no standardized vocabulary to use there, you don’t really need that. Most notes that have to do with accessibility, i.e. closed captioning or signing notes, could be included in 532 notes.

Is it recommended to use 538 in addition to 34X fields?

Yes, the 34x fields may not have comparable displays in local systems. Field 538 puts it into a form where people can read the note if that notes displays in your discovery system.

How do you describe multiple alphabets in field 546?

If you have multiple scripts, you can enter the scripts in separate subfields b, which is repeatable so you can enter multiple scripts in the note. On the slides there is an example with Mongolian in $a and Cyrillic alphabet in $b.

I see more and more books records using field 588. Is that correct?

You are probably seeing that more on electronic book records. This is required to tell you where the description came from for the e-book, for example the print book was used as the basis of the description for the e-book record.

Can you talk about the order the fields should be entered in?

There is not a prescribed order for notes in RDA as there is for AACR2 and for CONSER records that are in tag order. Catalogers tend to generally continue to use the prescribed order from AACR2.

Do you also use 588 field for 'Title from cover' note?

No, field 588 is about the whole description of the record. The source of the title would be entered in field 500. However, for continuing resources, field 588 will specify the volume that the description of the record is based on and will also include the source of the title.

What is the rationale for not transferring the various fields? It may be duplicative? For example, I found a GPO record that had been merged that lost the 500 field with the shipping list information.

The major rationale for not transferring certain fields in certain situations is that you may end of up with duplicative information. Since field 500 is for general notes, there is no telling what information may be in it, so there is no way of telling what may be important to retain. For manual merges, we are able to determine what would be important to transfer manually to the retained record. If you have found that information has been lost, you can add that back yourself, or if you are unable to do so, can send us a request to add that information back in. Send an email to bibchange@oclc.org, or use the error reporting function in Connexion and Record Manager.

Is it necessary to keep romanization parallel field, which is seldomly searched? Can catalogers just use the original language in item and ignore "Contents in vernacular field only" in 500, 520, 588, even in 505?

If a library has entered both non-Latin script fields and Romanized parallel fields for notes, please leave them in the
WorldCat record. If you wish to delete one or the other for your local catalog, assuming you do not use WMS, that is a local policy decision.

We have some old archival collection records that have 545 fields for biographical information of collection, organization, or person. Is this correct?

What you have sounds fine. Use of this field to provide more information for archival records is great. It is a field that would not be used routinely with modern published works.

For content notes input in volume number order in 505 fields, why is the order changed in reverse in OCLC (see OCLC# 231619396)?

While fields with non-Latin script are sorted ahead of fields with the same tag containing Latin script only, the issue here concerns the relative order of three 505 fields all of which contain non-Latin script. Reformating, validating, and replacing has no impact on the order of these fields as input by the cataloger. In the cited record we have moved the fields into their proper sequence and replaced the record in the Connexion client with no issues.

Is there a limit to the number of characters in the 505 specifically in Connexion?

Connexion Help says that record size in bibliographic records must meet size limits defined in MARC21 standards. The number of characters in a field cannot exceed 9,999. The number of characters in a record cannot exceed 99,999. These limits apply to records you catalog using Connexion and to those provided by the OCLC MARC Subscription service. For other offline services that output records and for catalog card production, record size is restricted to 50 variable fields and 4096 characters. Records may be truncated for the output records only. Connexion retains the full-length record. For record export, maximum record size is 6,144 characters, according to the online OCLC-MARC Records.

Is it ok to add the institution code in field 500 subfield $5 for non-rare materials?

No. If you are cataloging rare and special collection materials, you can use field 500, subfield $5 for copy- or institution-specific notes having scholarly or artistic value beyond the local institution. Please reference BFAS chapter 3.4.1 for more information.

For Chinese rare books, 500 fields are input in order by its cataloging rule, but after saving in OCLC, the order changes. How do you decide the 500 order?

You may be observing changes in the sort order based on fields containing non-Latin script always sorting ahead of those with Latin script only when they have the same tag.

Weren't rules changed so that both seen and unseen narrators are put in this field and not unseen narrators only in 508?

You are correct. It was formally rescinded with some simplification in RDA in the April 2015 Update. The instructions regarding Statements of Responsibility were greatly simplified, with much more being left to cataloger’s judgment. This is mostly thanks to a joint CC:DA task group of OLAC and MLA that tried to rationalize some complex instructions in RDA 2.4 (Statement of Responsibility), RDA 7.23 (Performer, Narrator, and/or Presenter), and RDA 7.24 (Artistic and/or Technical Credit). The instructions in RDA 7.23 and 7.24 were essentially deprecated in favor of references back to RDA 2.4 and 2.17.3 for Statements of Responsibility and forward to RDA Chapters 19 and 20 for...
“recording relationships to agents associated with a work or expression.”

I often think that a lot of 5xx fields require doubling in text what I would already be coding or describing in 3xx or 7xx fields. Do others recognize this? Is there a solution to not having to double the information?

For better or worse, MARC has always had built-in redundancies. The 007 fields code many elements that have been spelled out elsewhere, for example. That’s become even worse with the proliferation of 34x and other fields under RDA. In this transitional period at least, some local systems are not equipped to do anything useful with 34xs, for instance, so such fields as 538 remain useful in that sense.

Notes do not necessarily have to be included just to express information. Notes are no longer needed to justify access points in 7xx when they are not mentioned elsewhere in the description. But inclusion of these kinds of notes is not necessarily incorrect. Perspectives on this issue vary and are likely driven by what local systems display.

When inputting multiple 505 fields, should subfield $8 be used to indicate proper sequence? This seems logical but I don't think I've ever seen it.

Subfield 8 should not be used for these purposes.

Where is the best place to add a translation note, field 546 or field 500?

A note may not be needed if information about the translation or translator is transcribed in the 245 $c. If a note is needed, usually a 500 note is used. If there is complex information that involves both the language and the translation, a 546 field could be used. Examples:

500  $a Translated into English by Melissa Stone.
546  $a Original text in English with translations by Melissa Stone into French, Spanish, and Italian included.

Concerning field 546 vs. field 532--what if ASL is more of a main language than an accessibility feature? Recording of an interview carried out in ASL, for example.

When signing is the chief (or only) means of communication in a resource rather than an alternative accessibility feature, it would make sense to me to indicate this in field 546. It’s my hope that once we get some official guidance on the two recent accessibility fields 341 and 532, we’ll also have a better sense of how they are intended to relate with field 546.

In an email I received from Jay Weitz in January, he responded to the 588 note stating that it may be used for all kinds of resources. The newly revised PCC BIBCO Standard Record (BSR) RDA Metadata Application Profile, for instance, specifies on page 11 (corresponding to RDA 2.17.2) that the “Note on title” is a PCC Core element, when applicable. When the cataloger supplies the title, regardless of the type of resource, use of field 588 would be appropriate. Both the Objects Best Practices (page 30) and the BSR acknowledge that some institutions will continue to use field 500 for such notes because their local systems may not deal fully with field 588. But use of field 588 should now be the standard.

Although we most commonly associate field 588 with continuing resources, the field may be used for any appropriate type of resource. In addition to the BIBCO Standard Record Document cited, the current version of the Provider-Neutral E-Resource MARC Record Guide: P-N/RDA version cites two RDA elements that may use field 588: 2.17.13, Note on issue, part, or iteration used as the basis for identification of the resource; and 2.17.2.3, Title source. The latter instruction makes clear that it refers to a wide range of title sources, from print title pages to title frames of
moving images and has several examples backing that up. (In my reading of all of this, field 588 for a “Title from cover” note is fine.) As with many newer fields, some local systems may not be equipped to fully utilize field 588, so continuing to use field 500 is permissible but field 588 would now be preferred. Depending upon the circumstances and the significance of the title, a field 246 with an appropriate Second Indicator or subfield $i$ with display text may be useful to identify the source of a title.

Can vernacular scripts be added in 500 field and not paired with parallel Romanized fields?

Yes, though that is not standard practice currently within the U.S. If your language of cataloging is English, presumably your notes would be in English using Latin script. However, if you are quoting from the item that is in a non-Latin script, it is perfectly acceptable to include the non-Latin script in the quoted note. If your language of cataloging is Arabic, or another language using non-Latin script, then presumably your notes would be in that language and script.

You mentioned putting the source of title in a 500 field for a mono instead of 588. Is it incorrect to state in a 588 something like “Online resource; title from pdf title screen (Idaho Geological Survey, viewed November 5, 2020)”? Would it be better to include the title info in a 500?

Similarly, as I read all of this including RDA 2.17.2.3 and 2.17.13.4, some wording such as “Title from PDF cover page … based on version consulted: Nov. 5, 2020” is also perfectly acceptable in 588 or in 500 (as noted above).

November 12, 2020

What does it mean when a field is described as Required if applicable/Optional?

What we have in BFAS the standard for full followed by a slash, then the standard for minimal level. So, Required if applicable/Optional means that it’s required for full level and it’s optional for minimal level.

What is meant by "these fields don’t transfer when records are merged?"

In the merging process, whether it’s by DDR which is the automated process that runs through WorldCat or when we manually merge records, there are fields that will transfer if the retained record does not have that field, for example field 504. There are other fields that do not transfer, for example field 502, so we would have to manually transfer it when merging.

For the 504 field should we use webliography now instead of Includes bibliographical references?

If it presents itself as a webliography in the resource or a discography, then it’s okay to use that in field 504. Generally, if any kind of bibliographical chapter or appendix to a resource has a specific title, it’s okay to use the 504 note and to transcribe that title followed by a colon and the paging of the bibliography. If it’s just a standard bibliography, you want to follow the standard “Includes bibliographical references” followed by the parenthetical paging.

Concerning "does not transfer when records are merged"--is the field deleted from all the records, or if one record is considered the "main" record, the field would be retained, but not added from the other records?

In the process of merging, if the field is already on the retained record, that field will be kept. If the note is not on the retained record, it will transfer from the record being merged into the retained. If there are multiple records with the
If these are required fields, then shouldn't they be retained?

When we are merging manually and know a note is not going to transfer, we will manually transfer the note. We have more control over what gets transferred. In DDR, that’s an automated process with a complex set of algorithms around the transfer of data. We try not to lose important information but also at the same time trying not to add redundant information. For example, we do not transfer the 500 field because it’s a general note and we don’t know what kind of information may be in the note.

In the 518 example, shouldn't $a be $3?

Yes, you are correct! We will correct the typo before we post the slides.

I would be interested in standards regarding note order.

RDA does not specify a note order but AACR2 did, so it depends on what standards you are using for cataloging. The general practice is that you order the notes in order of importance. CONSER records are in tag order with the exception of fields 533 and 539 which are listed last.

Is it incorrect to put "Includes index" in the 505 $a?

A note that simply says “Includes index” should be entered in field 500. If it’s combined with a bibliographical references note, that could be part of the 504 note, i.e. Includes bibliographical references and index.

Someone said something regarding brackets around pagination in 504, I believe. Could you repeat that? Brackets used to be used for pages that didn't have a number printed on the page. But that seems to have gone away.

In RDA you wouldn’t use brackets.

Why are 500 notes indicating a video recording is in widescreen format being added to DVD/Blu-ray records when that information is already in the 538 field?

Field 538 does transfer, and if there is a record with this information is already in field 500, you end up with duplicate information.

The Marc Advisory Committee is in the process of defining a new subfield in an existing field for aspect ratio which would include things like widescreen and full screen, so that information in the future will have its own place in a MARC record, which it doesn’t have presently. That is why you often see aspect ratio information in field 500 because there was not a specific field for this kind of information. It’s also possible that a statement of aspect ratio may be properly included as an edition statement. That’s why you may see that kind of redundancy.

In the 502 field, do we omit "Thesis"? I did not see the 502 $a presented.

If you are using subfield $a to include all of the information about the thesis, then yes you would typically include “Thesis”. The example on the slide was a 502 with multiple subfields, so in that case you do not include the word Thesis.

Should the 538 note “Mode of access world wide web” not be added for e-resources since this would be...
obvious?

Yes, you do not need to add that note.

Regarding numbering, when is unnumbered used?

In field 300 when you are describing page sequences, you would say “X unnumbered pages”. If you need to specify which sequence you are talking about for the location of bibliographic references, you could give that in the parenthetical note. There is also the instance of where footnotes are scattered throughout the book at the bottom of the page, those may be bibliographical references you handle the same way.

Do OCLC staff have a sense of the extent to which information that should go in a specific 5XX field is input in a 500 field?

We don’t have any solid data on that, except gut reactions. We also have to remember that practices have changed and the MARC format has changed so that nowadays there are many more specific 5xx fields for which information that previously had been relegated to a 500 field would now be put in a specific 5xx field. One example, the aspect ratio that I mentioned earlier once that new subfield is defined.

Do you put “Dissertation” in the 502?

See answer above about “Thesis”.

Is a formatted 502 now required? The lack of punctuation in it would create an unpleasant display in a non-MARC environment.

No, it is not required. It depends on your system, so if your system does not take advantage of the subfields, then you may want to leave it all in $a. We have however been making a concerted effort to convert the 502 fields to the sub-fielded version. The reason for that is the information is much more granular and searchable.

So how is DDR (automated process) normally performed?

DDR is running constantly. Records that have been added as new records or records that have been changed all get fed every day into the DDR process with a delay of 7 days.

Is there a place where information is given on whether any given note field displays? We are a WMS/Discovery library and I noticed that 546 fields are not displaying. Other 5xx fields also don’t seem to display.

https://help.oclc.org/Discovery_and___a_is_displayed

If institutions are looking for ways to voice their opinion on what fields they want to see added to Discovery, then they can add something to the community center. When we’re ready to add additional fields to be displayed we always like to consult which ones are the most requested by the community.