2018-02-28 Virtual AskQC Office Hours

Topic presentation

Cataloging Defensively with Edition Statements, presented by Jay Weitz

Jay Weitz (Senior Consulting Database Specialist) presented on Cataloging Defensively with Edition Statements. The edition statement is a field that is heavily used in OCLC's matching and merging algorithms and through all formats. When to Input a New Record from OCLC Bibliographic Formats and Standards (BFAS), Chapter 4 has long served to provide a common basis for decision making in the creation of WorldCat bibliographic records by the participants in the OCLC Cooperative. When to Input has also been the public reflection of how OCLC's matching algorithms, including Duplicate Detection and Resolution (DDR), and automated loading of records are intended to work. When to Input has been thoroughly revised, with the updated version being available since October 2017.

One of the most important MARC elements that is included in the matching algorithms of both DDR and Data Sync is the edition statement. If a resource has an edition statement explicitly on the item, we ask that it be recorded in a field 250. An edition statement coded in a 250, rather than being recorded in a quoted or unquoted note or to other title information, can be one of the most effective means of differentiating records that might otherwise be identical. RDA 2.5.2.1 defines the scope of an edition statement including the words or statements that evidence a designation of edition. Other matching elements, such as dates or publisher numbers, may also adequately distinguish the records; however, an edition statement should be recorded in field 250 if one is present on your resource. In cases where there is no explicit edition statement on the resource, both RDA 2.5.1.4 and AACR2 1.2B4 give considerable power to catalogers to account for differences by supplying an edition statement. **Supplying an edition statement to bring out subtle differences in resources is one of the most effective ways of assuring that bibliographic records won't be merged incorrectly by DDR**. If there is a date associated with these different versions, it is within the scope of the instructions to include the date as part of the edition statement in field 250.

Member questions
Edition Statements

If a book lacks a first edition statement, but is otherwise identical, should you use the record?

Answer: Bibliographically speaking, a first edition statement is essentially ignored in the cataloging that we do every day. The lack of a first edition statement and a first edition statement are equal as far as determining whether to input a new record. If there is a bibliographic record that has first edition, but your resource does not say that, and if you have the option, you can edit the record locally if you want to remove that statement; however, a new record should not be created.

Why is a "Book club edition" statement in field 250 not a good justification for a new record even if that is the only difference?

Answer: Long-standing practice as outlined in When to Input a New Record takes "Book club edition" and says not to create a new record if that is the only difference. There will be cases where a book club edition has other differences (paging, size, etc.) that would have an impact on the content of the item so that page for page it's not the same. When you look at book club editions, in many cases, they are just a cheaper binding and that page for

page it really is the same content. That was the original decision that went into putting that criteria into When to Input a New Record, to say discount book club editions.

I have a Spanish language book with several edition statements. The only difference between the statements is the month of publication. Are these "real" edition statements, or just printings?

Answer: This is what we have come to incorrectly or to narrowly refer to as the Romance language problem. In certain language such as Spanish, Italian, German, and various other languages what look to be edition statements but are printing statements and are often associated with a number of copies that are printed. These should be generally ignored and not even transcribed, although practices have differed over the years. Those are not considered to be real edition statements. They are just printing statements and should not be a factor whether you input a new record. They should be ignored.

Is it preferable to have two separate 250 fields like in the case of your example "Content 3rd ed. Pew ed. with readings" rather than recording all the info in one field? For example: "3rd ed., Pew ed., with readings (all in one field)?

Answer: That will depend on the resource. There are some things, particularly with scores, where what used to be separate edition statements versus what used to be called, under AACR2, musical presentation statements. Where the statement in the edition statement and the statement that was in the 254 field, are grammatically and intellectually separate. So those would be legitimately separate 250 fields. It is a matter of judgement. Field 250 was made repeatable just a few years ago to accommodate multiple edition statements, where previously we were obligated to record it in a single 250 field, separating one edition statement from another by a comma. That no longer must be the case.

I catalog a great many state government publications. It is not uncommon to have two or more versions of a document published in the same year. The physical descriptions are identical and there are no edition statements. The only difference between one and another is something textual, such as an updated directory of contacts or the updated text of a law. Over the years, I have been frustrated by having some of my document records merged because I described the textual differences on in a note field. Thank you for your suggestions for devising edition statements.

Comment: If you come across records that have been merged incorrectly, please report them to us. Depending on how long ago they were merged, they can often be recovered. In cases where we can recover an incorrect merge, we or you can often supply something that will help differentiate the records in the future so that DDR won't incorrectly merge them again. In government documents, it is very common to have multiple documents published in the same year. If you can identify a particular month or date in addition to the year and include that information in a supplied edition statement, that is one option to prevent the records from being merged. DDR tries to look for certain quoted information, either specific publication numbers or serial numbers, that may appear in quoted 500 notes.

If a 250 field is associated with a subfield $3, does that affect the field's value for distinguishing one version from another?

Answer: Subfield $3 in a 250 field is not taken into consideration in DDR.

How does OCLC handle a record submitted with one language in the 008 field and a different language in the 250 field?

Answer: If the error is reported to us, we will fix it. There is no automatic mechanism to alert us about something like that.

Are galley copies considered different editions?

Answer: Yes, galley copies are prepublications, different editions. So, if you have a galley version, you can include that in a 250 field and DDR will not merge that to the published version or other similar things. Many of the books that we pick up at ALA that are advance reader's copies, you can use "Advance reader's copy" in a supplied edition statement, or if that is what is says on the item you can use that in the edition statement.

General questions
Notes Fields

Is the 502 Dissertation Note field solely used for unpublished dissertations? Can this field also be used for published dissertations, or should we use field 500 for published dissertations?

Answer: In the United States, dissertations in 502 fields should be used only for the actual unpublished dissertation. References to the fact that it is an adaptation of a dissertation, or along those lines, would be recorded in a 500 note. DDR does try to make that distinction between the published version and unpublished version of a dissertation, including looking for the presence of field 502 in addition to other elements as well.

The definition of field 501 says "A note indicating that more than one bibliographical work is contained in the physical item at the time of publication, release, issue, or execution. Use field 501 only when the physical item is not being described as a unit. The works that are contained in the item usually have distinctive titles and lack a collective title. Field 501 usually begins with the designation Issued with:, etc.". Does it mean that the different parts of the physical item are published at the same time, or can they be published at different times? Can we still use field 501 even if the different parts of the item (bound in one volume) are published at different times?

Answer: The 501 field (With note) is generally not used in RDA. It should be used properly only when the physical item is published as that conglomeration of things, not for things that your institution has bound together after their individual publications.

BIBFRAME

I've been hearing that when BIBFRAME is eventually implemented that MARC will become obsolete. Is that true and if so, will any aspects of MARC still be valid in the future?

Answer: Cannot definitively say anything for sure; however, there will be no one day at a point where we will map all of our MARC data to BIBFRAME and everyone will all switch over to use BIBFRAME exclusively. There will be a long time period where data is created in one format and needs to be mapped to the other. We will probably see a lot of back and forth, moving data around from one format to another. Even though MARC may be on its way out in the long run, it will probably be in use for a good number of years.

Series Statements

Recently I have seen a lot of Library of Congress bibliographic records with series statements that have no authority records to match them. Indeed, sometimes I see different forms of the series in an 800 and an 830 field. Is there any policy that series must have a matching traced series statement to be added to an LC record, or does it depend on if it is a PCC record or not?

Answer: The Library of Congress made a decision a number of years ago, widely known as the "LC Series Decision", where they decided as an individual library that they were no longer going to make series authority records. The PCC continues to make series authority records in PCC libraries outside of LC, but there is no requirement within PCC that a series authority be made. The idea being that if you do have a traced series, meaning that you are going to use it as an access point and have an 8XX, you do want to have a series authority record if you are a PCC library. LC just transcribes the series in a 490 field and does not add an access point in an 8XX field.

Dates

If all elements are identical, including ISBN, but dates are different (one year apart) should you use the record?

Answer: If it's what we have come to call a "trade publication" (a major publisher), you may want to use an existing record even if the date is a year off. That depends on the individual instance and your own judgement.

DDR

Does DDR also look at the Date type to differentiate a date such as 1999 03 vs. 1999 04 in the fixed field?

Answer: No. In theory, if that information is reflected elsewhere in a record such as the edition statement or possibly in a quoted note, DDR would pay attention to that.

If I enclose a date in quotes in a 500 field, would this prevent an improper merge?

Answer: We try to identify dates in quoted notes. There are lots of different ways to present a date (mmddyy, yyyymmdd, etc.) and which part of the date should come first, second, or third. We try to interpret and parse those different methods of quoting dates in 500 quoted notes. In many cases that will prevent an improper merge; however, if that information is in a 250 edition statement it is more easily interpreted and seen as a differentiating factor that will prevent an improper merge.

Why are there so many bibliographic records in OCLC with encoding level M when many of the records are the same or very similar?

Answer: Encoding level M indicates that the record was batchloaded without a human looking at the record trying to compare it to other records that are in the database, but our algorithms trying to find a duplicate and to merge it to that. Our algorithms are not perfectible, they are imperfect and don't have the advantages that we humans have of being able to interpret the information. We try to bring together records that should be together, but we also try to do our best to keep apart records that should be kept apart. It is a balancing act to merge records that should be merged and keep records part that should be kept apart. If you find records that are duplicates, report them to us. If you find records that have been merged incorrectly, report those to us as well. We learn something

from every incorrect merge, and many incorrect merges allow us to further finetune our bibliographic and matching algorithms.

RDA

We've noticed that RDA fields, like the 33X fields, have not been added to most of the older serial records. Are there plans to add those to serial records?

Answer: Yes. When we started the RDA hybridization of records to add fields like the 33X fields, and other things like the spelling out of abbreviations in field 300, we started with Books because they are the easiest ones to deal with and then went on to other formats. Serials still need to be done. One of the factors we've considered is how many of the CONSER records will be affected, because the changes that we make are then transmitted in the file that we send to LC. We will continue to make headway on Serials and all of the other records that are in the database that don't have 33X fields.

Reporting Incorrect Merges

Who would I contact about incorrect merges? How do I contact that person?

Answer: Incorrect merges can be sent to bibchange@oclc.org.

Errors Messages

Why am I given an error on a particular field when I try to add my holdings and I did not add that field to the record, for instance the 084 field?

Answer: There is a way to change your validation level check in settings. You may have it set to where it is doing a full validation. That is something that you can minimize.

77X References

Any plans to create a process to delete 77X references to cancelled records? When cleaning up electronic records, I am deleting a lot of these. A related question, is OCLC able to search and attach to existing records for electronic titles vs. making duplicates when they are generating HathiTrust records?

Answer: The process that creates HathiTrust records is out of date and needs work. We have put together requirements for what needs to be done with that, but we do not have a timeline to when that will take place. We have certainly thought about a process to delete 77X references, but we don't have a mechanism in place that can go through the database and detect that those fields exist and that they are pointing to records that no longer exist. It is something that we will keep in mind and try to get something in place if it is possible to do so.

Does having a 776 field in the print help prevent duplicate the HathiTrust duplicates?

Answer: No, not right now.

006 Field

I am now finding the 006 a e 000 1 in some records for CD audiobook sound recordings. This field for "Additional Material Characteristics" appears to be coded for "Books" "Adult" Fiction" however these CDs do not appear to have additional characteristics, there is no CD-ROM or pdf added material. Should this 006 field be retained in bibliographic records for Sound Recordings when the resource cataloged is just a spoken word sound recording?

Answer: It would seem that the 006 field does not belong on those types of records. If you are unsure as to whether or not to remove those fields, you can send those to bibchange@oclc.org for further investigation.

Type of Record

In Bibliographic Formats and Standards, the Type of Record (Leader/06) code for mixed materials (code 'p') says that "for made-up collections in which one form of material predominates, use the appropriate code for that predominate material". A search in WorldCat for archival materials only looks for that 'p' code. What about archival material that is purely textual, or any other single format?

Answer: In WorldShare and WorldCat Discovery, the Material Type "mix" does retrieve only records coded as Type of Record (Leader/06) "p", according to *Searching WorldCat Indexes*.

"When performing a command-line search in Connexion or an expert search in FirstSearch, WorldShare, or WorldCat Discovery" the Material Type search "mt:arc" (for Archival Material) should retrieve all records coded as "a" (Archival) in Type of Control (Leader/08) (according to *Searching WorldCat Indexes*. Note the difference between Type of **Record** (Leader/**06**) and Type of **Control** (Leader/**08**).
Although this additional fact is no longer accurately reflected in *Indexing*, as far as I can tell, the Connexion indexing of "mt:mix" should actually retrieve Type of Record (Leader/06) values p, t, d, and f. A few test searches in Connexion suggest to me that this is definitely still true.

Type of Record (Leader/06) value "p" should be used for "Collections of materials in two or more forms that are usually related by virtue of having been accumulated by or about a person or body. … This category includes archival and manuscript collections of mixed forms of materials such as text, photographs, and sound recordings." Manuscript and archival collections that are primarily textual should be coded Type of Record (Leader/06) value "t"; primarily cartographic should be coded "f"; and primarily notated music should be coded "d".. For all of these sorts of mixed/archival material collections, Type of Control (Leader/08, Ctrl) should be coded with value "a".

Future Office Hours Sessions

Will there be additional sessions like this webinar in the future? If so, where can I sign up?

Answer: We are planning to hold these on the last Wednesday of each month through the end of June this year. We will then evaluate and decide whether to continue beyond that point or not. You don't need to sign up any place, you can just use the login information and log in at 1:00 PM Eastern Standard Time. It will be the same login information that was used for today.

Will the recording and notes be added to the OCLC website? What about the login information with WebEx link and phone number to call in?

Answer: The recording will be posted for this session. We will announce this on OCLC-CAT when the recording is posted. We may also post some notes as well and will announce on OCLC-CAT when those are posted and where they're posted. Today's presentation will be added to the Cataloging Defensively page on the OCLC website as well. We can post the login information at the same place that we post the recording. This information will also be announced on OCLC-CAT and included in the Message of the Day in the Connexion login a few times before each session.

Are you looking for suggested topics for future sessions?

Answer: Yes, please send your ideas in to askqc@oclc.org.