Virtual AskQC Office Hours

The complexities of field transfer

OCLC Metadata Quality

September 2021

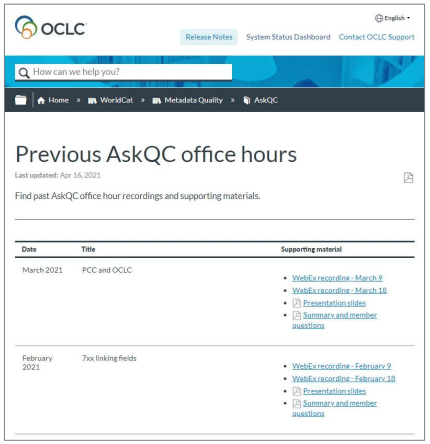
Housekeeping

This session is being recorded

Housekeeping

This session is being recorded

All session recordings, slides, and notes are available at oc.lc/askqc

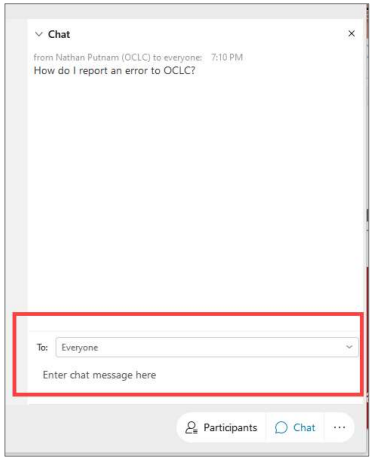


Housekeeping

This session is being recorded

All session recordings, slides, and notes are available at oc.lc/askqc

Enter questions in chat to "Everyone" at any time during the presentation



Housekeeping

This session is being recorded

All session recordings, slides, and notes are available at oc.lc/askqc

Enter questions in chat to "Everyone" at any time during the presentation.

After the session, you will be directed to a brief, optional survey

On the call today

Charlene Morrison
Database Specialist

Laura Ramsey
Senior Metadata Operations Manager

Shanna Griffith
Database Specialist

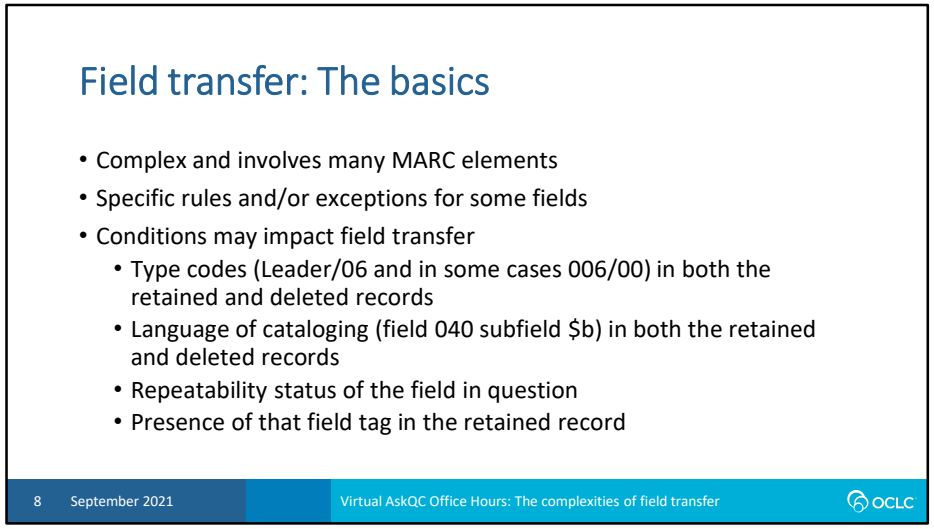
Robert Bremer
Senior Consulting Database Specialist

Bryan Baldus
Consulting Database Specialist

Hello. Today Laura and I are going to talk about the complexities of field transfer.

For those who may not be familiar, field transfer is the selected bibliographic data from a record identified as a duplicate that are copied (or transferred) into the record identified as the retained record. Software performs field-by-field transfers according to strict and complex rules. We will take a closer look at some of those finer details during this presentation.

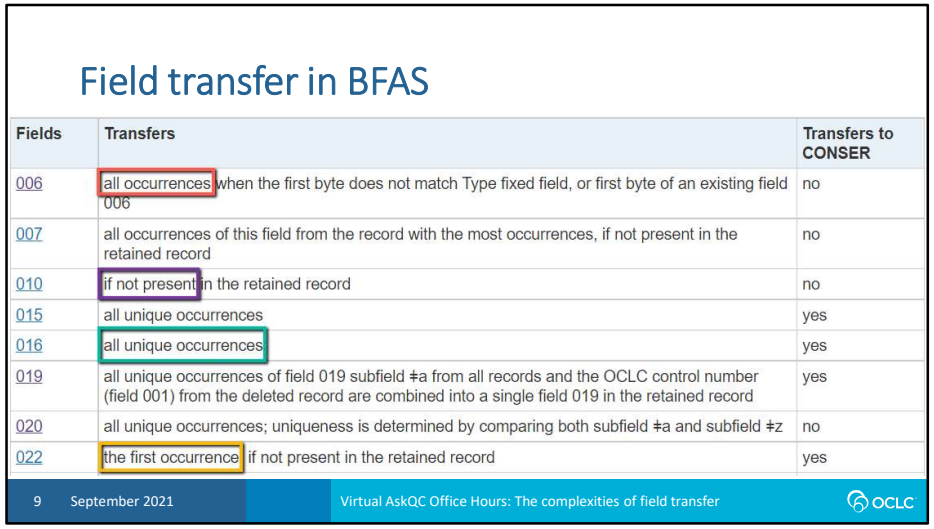
I will be covering some of the details around field transfer for manual merges and Duplicate Detection and Resolution (AKA DDR). Later, Laura will highlight some of the details of field transfer for other processes, such as Data Sync.



Field transfer is complex and takes many MARC elements into consideration. There are several standard sets of transfer rules; however, some fields have their own specific rules and/or exceptions. Although records that are candidates for deduplication may be within a set of any size, in practice, they are compared in a succession of pairs. Meaning within the set of records, one is considered the retained record and the others will be merged and deleted.

There are some conditions that may have an impact on field transfer, such as:

- Type codes in both the retained and deleted records
- Language of cataloging in both the retained and deleted records
- Repeatability status of the field in question, and the
- Presence of that field tag in the retained record, just to name a few



Field transfer in BFAS

Fields	Transfers	Transfers to CONSER
006	all occurrences when the first byte does not match Type fixed field, or first byte of an existing field 006	no
007	all occurrences of this field from the record with the most occurrences, if not present in the retained record	no
010	if not present in the retained record	no
015	all unique occurrences	yes
016	all unique occurrences	yes
019	all unique occurrences of field 019 subfield ǂa from all records and the OCLC control number (field 001) from the deleted record are combined into a single field 019 in the retained record	yes
020	all unique occurrences; uniqueness is determined by comparing both subfield ǂa and subfield ǂz	no
022	the first occurrence if not present in the retained record	yes

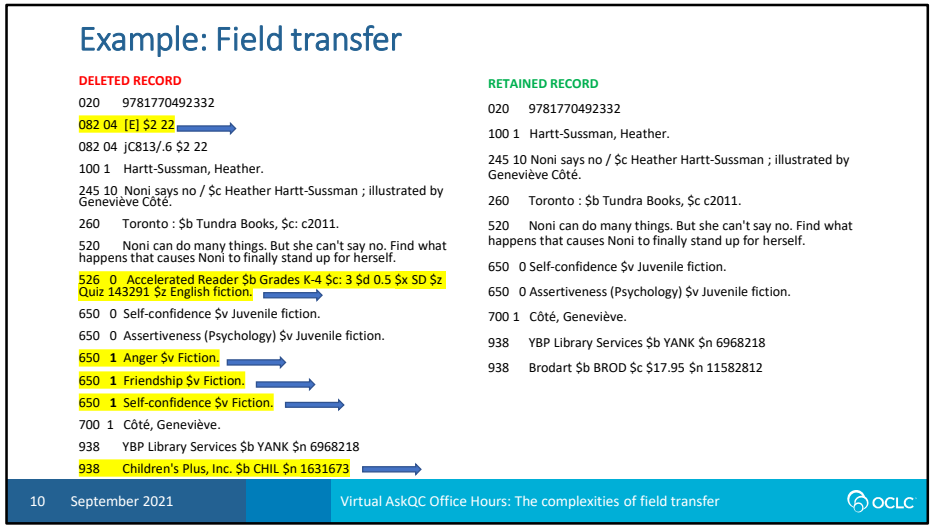
In Bibliographic Formats and Standards (AKA, BFAS), we have added a Field Transfer section to Chapter 5.4 WorldCat Metadata Quality Activities. This table provides insight to what happens through the DDR and manual merge processes. It provides a list of the fields that can transfer, the conditions that must be met for the field to transfer, and if that field will transfer to CONSER records. CONSER records are continuing resource bibliographic records contributed by the Cooperative Serials Program of the PCC and can be identified by the presence of an authentication code such as pcc in field 042.

Some of the terms that you will see repeated throughout the table are outlined in the various colored boxes on the slide.

- All occurrences: is every instance of the field in a set of records
- If not present: means the field does not exist in the retained record
- All unique occurrences: is every distinct instance of the field in a set of records that meets the specific criteria for that field, and
- First occurrence: is the initial instance of the field in a set of records

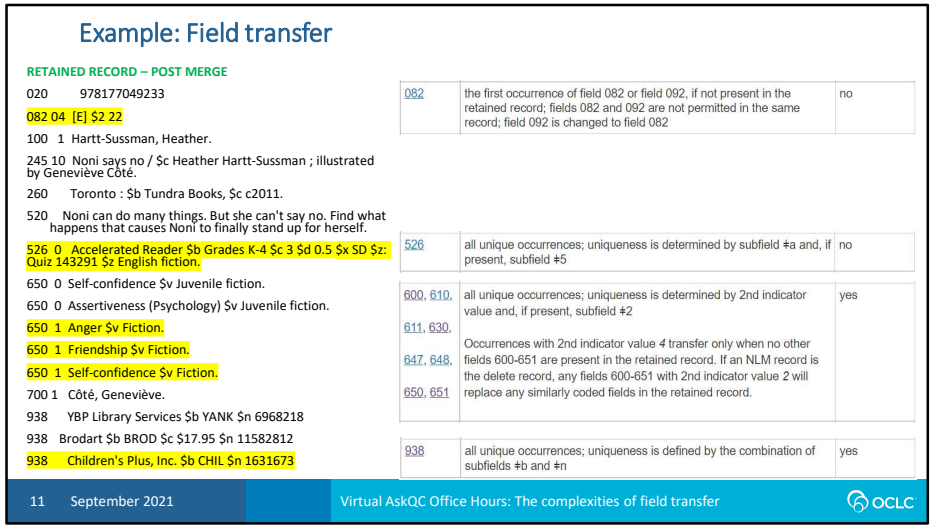
So, in this section of the transfer table on the slide, we can see that:

- field 010 (Library of Congress Control Number) will transfer if that field is not present in the retained record and is marked ‘no’ it will not transfer to CONSER records
- all unique occurrences of field 016 (National Bibliographic Agency Control Number) will transfer to the retained record and is marked ‘yes’ it will also transfer to CONSER records
- And finally, we see that the first occurrence of field 022 (International Standard Serial Number) will transfer if the field is not present in the retained record and is marked ‘yes’ it will also transfer to CONSER records



Here we have an example of duplicate records. The duplicate (or delete) record on the left has fields that are not present in the retained record on the right and will transfer during the merge process, if the transfer criteria for each field is met.

Note that only one of the 082 fields is highlighted as a field to transfer, and the second indicator has been bolded in the highlighted subject access points.



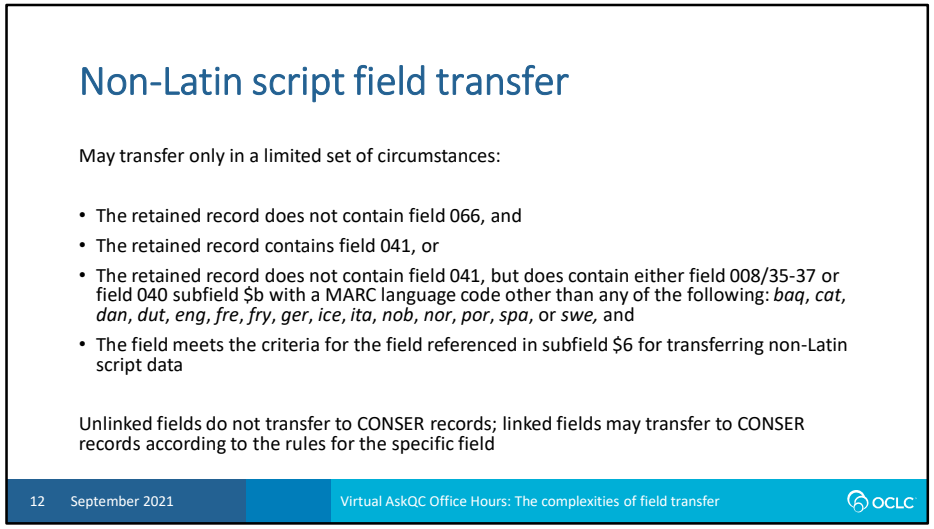
Here is the image of the retained record after the merge.

We can see from the transfer guidelines on the right for each field, that only the first occurrence of field 082 will transfer, if not already present in the retained record.

The transfer guidelines for field 526 are for all unique occurrences, with uniqueness determined by subfield $a and, if present, subfield $5. In this case, since this field was not already present in the retained record, it is considered unique and was transferred.

For fields 600-651, uniqueness is determined by the 2nd indicator value and, if present, the subfield $2. Fields only transfer if no other subject access points in that same scheme are present in the retained record. So, these LC children's headings, with second indicator value 1 transferred, since none with that same second indicator value were already in the record.

And, field 938 transferred based on the uniqueness of subfield $b and subfield $n.



As previously mentioned, some fields have specific rules and/or exceptions that must be met for the field to transfer. The non-Latin script field 880 is one such field and may transfer only in a limited set of circumstances.

- The retained record does not contain field 066, and
- The retained record contains field 041, or
- The retained record does not contain field 041, but does contain either field 008/bytes 35-37 or field 040 subfield $b with a MARC language code other than any of the following codes shown on the slide, and
- The field meets the criteria for the field referenced in subfield $6 for transferring non-Latin script data

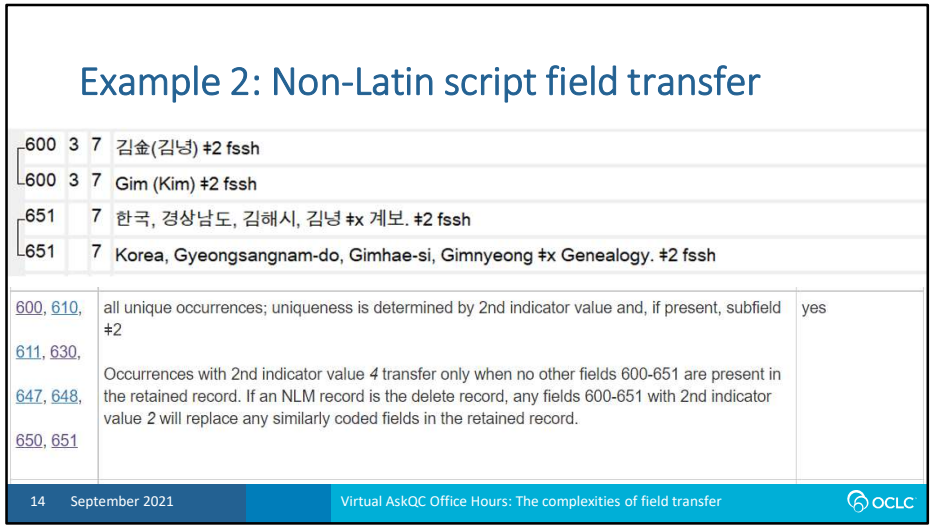
Unlinked fields do not transfer to CONSER records; linked fields may transfer to CONSER records according to the rules for the specific field

Example 1: Non-Latin script field transfer

520			本书以二十四节气为时间线,告诉读者每个节气的养生重点是什么,选什么应季菜更适合,内容包括:春吃鲜,夏吃味,秋吃韵,冬吃藏.
520			Ben shu yi er shi si jie qi wei shi jian xian,Gao su du zhe mei ge jie qi de yang sheng zhong dian shi shen me,Xuan shen me ying ji cai geng shi he,Nei rong bao kuo:chun chi xian,Xia chi wei,Qiu chi yun,Dong chi zang.

520	all occurrences, if not present in the retained record	no

In this first example, we have the 520 field. The transfer table shows that all occurrences of this field will transfer if not already present in the retained record. We also see that it is marked as ‘no’, it will not transfer to CONSER records.

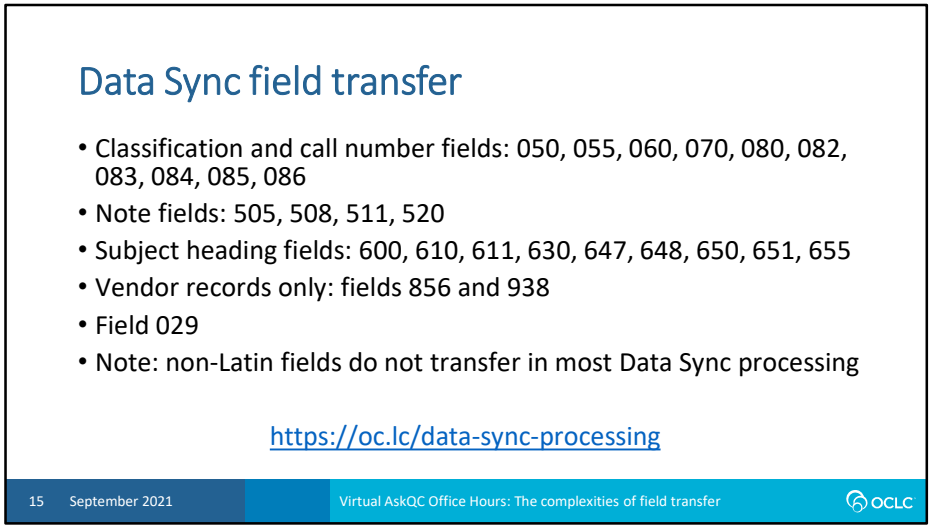


In this second example we have a couple of 6XX fields, both with second indicator *7*. The transfer table shows that for the listed fields 600-651, all unique occurrences will transfer to the retained record. Uniqueness is determined by 2nd indicator value and, if present, subfield $2. In this case these fields would transfer to the retained record if this scheme was not already present in the retained record.

The table also provides additional information about occurrences with 2nd indicator value *4*. Noting that they will transfer only when NO OTHER fields 600-651 are present in the retained record.

In addition, this section also shows how transfer of these fields are handled if the delete record is an NLM record.

I will now pass it over to Laura to talk about field transfer for other processes and changes to field transfer.



Shanna covered field transfer for DDR and manual merging. I'm going to cover the difference in field transfer for other processes. Most records that are processed through Data Sync or other batchload projects fall into this category of a limited set of fields that are eligible for field transfer. There are exceptions to this for certain projects, such as records contributed by national libraries or through a cooperative cataloging program such as PCC, which have the same criteria for fields eligible for transfer as DDR.

You'll see from this slide that we transfer fields in the categories of classification and call numbers, some selected note fields, and subject heading fields. These are all transferred using the same rules that are outlined in the field transfer table in chapter 5 that Shanna referred to.

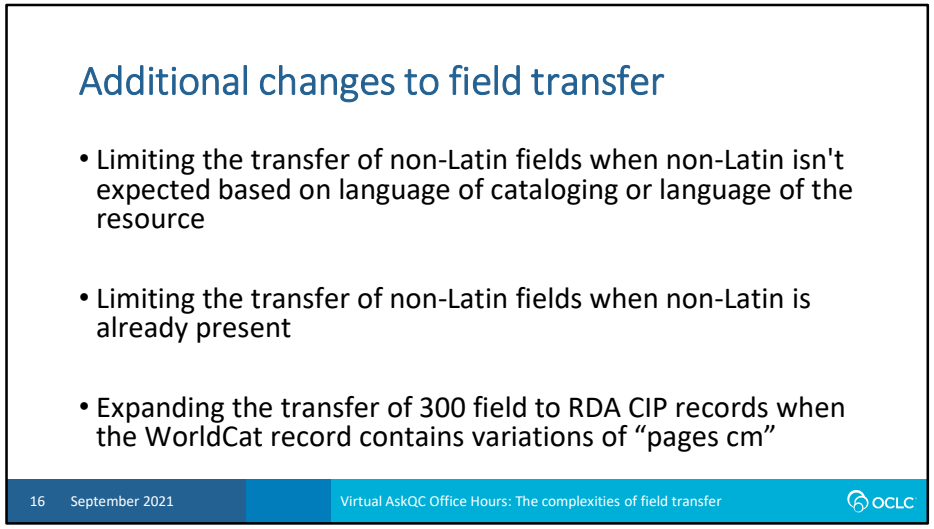
One exception to that is field 856 which will only transfer in vendor record processing and from national and cooperative libraries projects. In fact, several years ago we changed the criteria in order to reduce the transfer of 856 fields with local URLs, however you may still see remnants of that in WorldCat as we continue to eliminate local URLs.

The other exception to most projects is field 938, which is vendor specific information, so therefore only transfers in

vendor record processing.

And one last field that we transfer that doesn't fit into the categories just described is field 029. I also want to call out that non-Latin fields generally do not transfer in Data sync processing.

For more information on field transfer in Data Sync processing, see https://oc.lc/data-sync-processing.



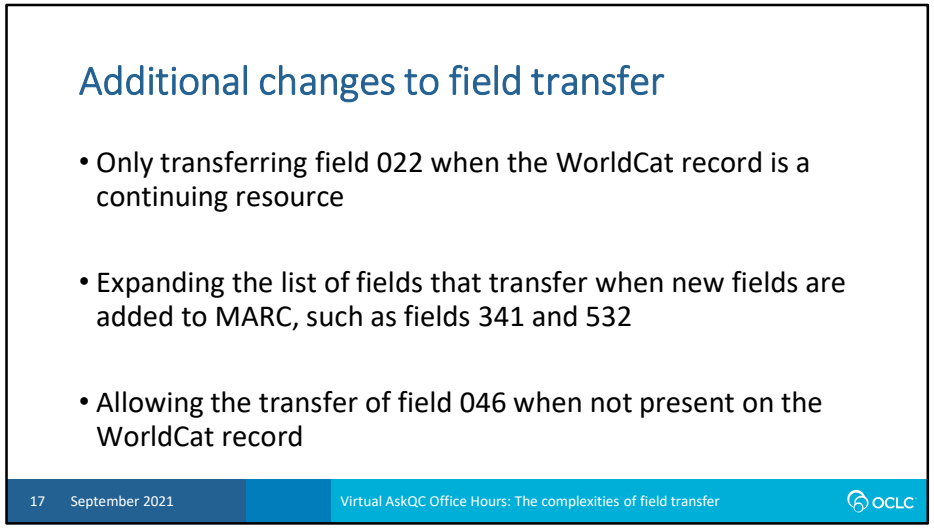
Now back to the overarching topic of field transfer. In addition to changing the transfer rules for field 856 as I mentioned, we are continually making adjustments to field transfer for reasons such as new MARC fields, unexpected outcomes of field transfer, and refinement or improvement of the rules.

All of these changes that I’m going to highlight apply to both DDR and Data Sync field transfer when the fields are candidates to transfer according to the type of processing.

Some recent changes made to field transfer involve the handling of non-Latin fields to limit the transfer when non-Latin field is not expected for the WorldCat record, as well as preventing the transfer if non-Latin is already present on the WorldCat record. Shanna went into more detail about the criteria for the transfer of non-Latin data.

We also expanded the transfer of field 300 to RDA CIP records when the 300 field on the WorldCat record contains

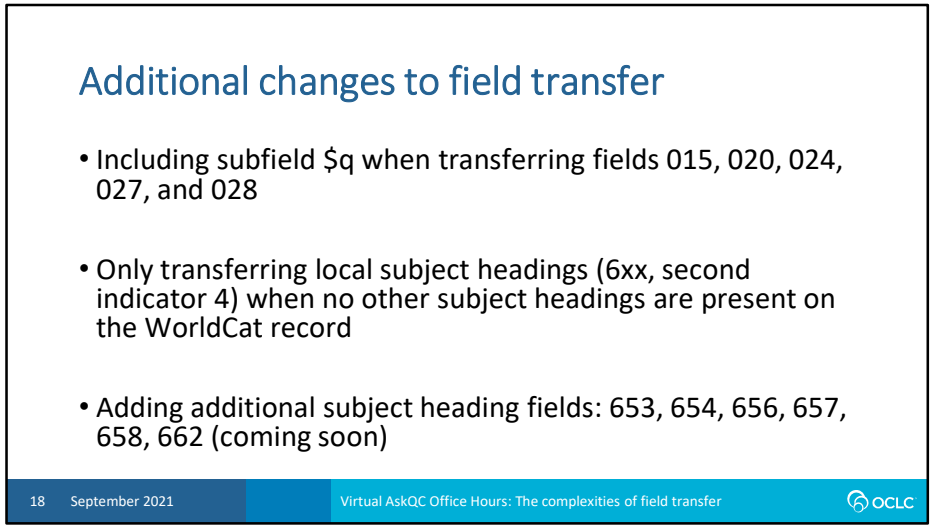
“pages cm”, where previously the rule was to transfer when it was “p cm”.



Other recent changes include only transferring field 022 when the record is a continuing resource. Field 022 is only valid on continuing resource records, and this eliminated the undesirable transfer to non-continuing resource records and the associated validation error when a cataloger attempted to replace the record.

We also consider expanding the list of fields whenever new fields are added to MARC. Recent examples of this are field 341 for accessibility content and field 532 for an accessibility note.

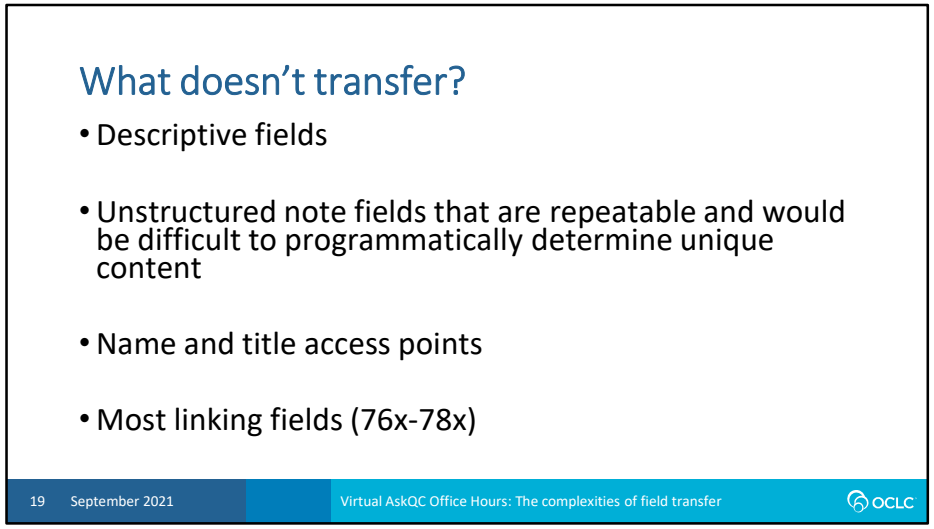
Field 046 was a recent addition as a field eligible for transfer.



We also changed the criteria to include subfield $q when transferring fields 015, 020, 024, 027, and 028.

A major improvement was made when we stopped the transfer of local subject headings when subject headings were already present on the WorldCat record. This eliminated the duplication of subject headings that gave many catalogers headaches.

And coming in the near future, we are expanding the list of subject heading fields to include fields 653-658 and field 662.



We often get questions as to why certain fields did not transfer when records were merged or matched to. We want to be sure to transfer data that will add value to the WorldCat record, but not duplicate existing data in the record or potentially change the identity of a record.

Descriptive fields are out of scope for field transfer since the description of a record should only be modified by a cataloger with the resource in hand.

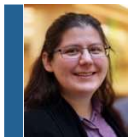
To avoid duplication, there are many notes fields that are unstructured that we do not transfer since it would be difficult to determine when the content of a field is already present in a note on the WorldCat record. One such example of this is field 500.

Also out of scope are name and title access points, which should also be cataloger supplied. And most linking fields are not eligible in order to avoid the potential of duplication when such fields have differences in content, and to also avoid potential incorrect references to other WorldCat records. Again, those fields should be cataloger supplied to ensure their accuracy.

If you suspect there has been a loss of valuable data when records are merged or matched to, we are always happy to look into it and add it to the WorldCat record if appropriate. Such requests can be sent to Bibchange@oclc.org or through the error reporting function in Connexion and Record Manager.

And now I'll turn it back over to …

On the call today



Charlene Morrison
Database Specialist

Laura Ramsey
Senior Metadata Operations Manager

Shanna Griffith
Database Specialist

Robert Bremer
Senior Consulting Database Specialist

Bryan Baldus
Consulting Database Specialist

Thank you!

October Virtual AskQC Office Hours
Getting a fix on fixed field elements

Tuesday, 12 Oct. at 10:00 AM Eastern
Thursday, 21 Oct. at 4:00 PM Eastern

Registration and session links available at oc.lc/askqc

Send cataloging policy questions at any time to askqc@oclc.org

Photo by Eric Rothermel on Unsplash