Virtual AskQC Office Hours
Languages, Non-Latin Scripts, and Mysterious MARC 880 fields
OCLC Metadata Quality
March 2023
Housekeeping

This session is being recorded
Housekeeping

This session is being recorded

All session recordings, slides, and notes are available at oclc/askqc
Housekeeping

This session is being recorded
All session recordings, slides, and notes are available at oc.fc/askfc
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.fc/askoc](http://oc.fc/askoc)
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  
Senior Data Analyst

Sharma Griffith  
Senior Data Analyst

Cynthia Whitacre  
Senior Marketing Operations Manager

Jay Weitz  
Senior Executing Database Specialist

Bryan Bakhus  
Senior Data Analyst
Thank you, Charlene. As you know, recording the languages associated with a resource can be somewhat confusing given the many different types of data and the areas within the bibliographic record this information can be found. During this presentation, Cynthia and I will cover some of those areas.
In particular, here are the topics and their corresponding fields that we plan to cover today. We’ll go in this order, which is not exactly MARC field order, but hopefully it will make sense.
The first is the language of the resource which is recorded in the fixed field Lang element. This is not to be confused with the language of cataloging that is recorded in field 040 subfield $b$, which Cynthia will talk about in a moment.
The fixed field Language Code element is a mandatory three-character code that represents the language of the resource and is applicable to all format types. While only one language may be coded in this element, I will show later how you would code for a resource that is multi-lingual.

The default code is three fill characters, meaning no attempt was made to code.

It is indexed in the Language Word and Language Phrase Indexes and is used in conjunction with field 041.
Field 041 is also the Language Code and contains the MARC language codes for the languages associated with a resource, such as the original language for a translation, librettos, sung or spoken text, etc. You do not use field 041 to record the MARC language code if there is only one language associated with the resource already coded in the fixed field.

The field is repeatable, is indexed in the Language Word Index, and is used in conjunction with the fixed field Lang element.
As you can see, field 041 has A LOT of subfields that can be used for recording a variety of types of languages associated with a resource. These next 2 slides show the subfields, whether they are repeatable or not, and those that were recently added over the last few years, such as subfield $i$ Language code of intertitles, which was added in 2019.
And here are the remaining subfields. Subfield $t$ Language code of accompanying transcripts for audiovisual materials was added in 2019, followed by subfield $7$ Data provenance in 2022.

Please refer to the page for this field in Bibliographic Formats and Standards for examples and more information.
As previously mentioned, the coded language information in the fixed field `Lang` element and in field 041 are used in conjunction when the fixed field element is insufficient to convey full information for multilingual resources, resources that are translations, or resources where a medium of communication is sign language. Field 546 is used to record the language information in textual form.

These examples show a couple common use cases between these elements – multilingual resources (or items in multiple languages) and translations.

The top example is for a multilingual resource that has the various language codes for the title, text, place names, and legend in separate field 041 subfields $a$. The 1st indicator is coded ‘0’ to indicate the item is not a translation or does not include a translation. For works in multiple languages, the codes for the languages should be recorded in the order of predominance. If predominance cannot be determined, record the codes in English alphabetical order.

The bottom is an example of an English translation of a book written in French. The 1st indicator is coded ‘1’ to indicate the resource is or includes a translation. In subfield $a$, the code of the language of the resource is recorded. In subfield $h$, the code of the original language is recorded.
For musical works, the language code corresponding to the textual contents should be recorded in subfield $a$, only when the work is printed or manuscript music. If the work is an audio recording, record the language code in subfield $d$.

The first example is the audio recording of a vocal work sung in Italian as recorded in the 041 subfield $d$, with accompanying program notes in English (recorded in subfield $g$) and libretto in Italian with English translation (recorded in subfield $e$).

The second example is the score of a vocal work which has lyrics in Italian as recorded in subfield $a$ and subfield $h$, with a German translation (recorded in subfield $a$) accompanied by an English preface (recorded in subfield $g$ and subfield $m$) with German translation (recorded in subfield $g$).

The second example is the score of a vocal work for which the libretto is in both Italian and German and preface in English with German translation.
While OCLC prefers the use of MARC codes as found in the MARC Code List for Languages, field 041 may be repeated if non-MARC codes are used, such as those from ISO. These codes may provide more granularity versus MARC, in that a language group may be under the MARC code versus the individual code in ISO.

In this example on the slide, the resource has text in Iduna translated from English. In the MARC Code List for Languages, Iduna is subsumed under the collective code ‘map’ for Austronesian (Other). In the ISO 639-3 list of Codes for the Representation of Names of Languages – Part 3: Alpha-3 Code for Comprehensive Coverage of Languages, Iduna has been assigned its own code ‘viv’.

I will now pass it over to Cynthia.
Thanks Shanna.

Language of cataloging is the language a cataloger uses to catalog the resource. It is NOT the language of the item, though they may be the same. One may catalog a resource that is in Swahili using English language of cataloging. One may catalog a book that is in English with Italian language of cataloging. Language of cataloging is what you use to describe the item in the 300, 33X, and non-quoted notes in 5XX fields.

Language of cataloging affects the language in which you record the physical description, the language of the 33X fields, and the language of notes. That is illustrated on this slide. If you are cataloging in Spanish, the fourth example, which is in the column on the right, illustrates using the Spanish word for pages, for volumes, and for index.

In the United States, most cataloging agencies use English as their language of cataloging. That means you will code 040 subfield $b using "eng", no matter what the language of the resource being cataloged. If you are cataloging the resource in French, which is likely for our cataloging colleagues in Quebec, you code the $b fre for French.

Note that some cataloging agencies DO use multiple languages of cataloging, but that is unusual and would be a local policy. For example, the National Library of Israel catalogs using different languages of cataloging depending on the language and script of the resource.
Bibliographic Formats & Standards outlines OCLC policy for Language of cataloging in 2.6. Here is the beginning of that text. OCLC calls records for the same resource in different languages of cataloging parallel records. Notice that second sentence which is highlighted. Only one record per language of descriptive cataloging is allowed within WorldCat for each manifestation of a work.

Language of cataloging is incorrect in thousands of records in WorldCat, since many mistakenly code 040 subfield $b$ using the language of the resource, rather than the language of cataloging. Coding this correctly is crucial for matching within WorldCat, Both DataSync, OCLC’s batchloading mechanism, and DDR, our duplicate detection and resolution program, only match and merge records with the same language of cataloging. So, if it is coded incorrectly, a duplicate record will very likely be the result. A real example of incorrect coding I encountered and fixed was a collection of sound recordings sung in Cajun French. They were all coded with language of cataloging as "roa", Romance Other, which was the correct coding for the language of the item. However, they were all cataloged in English, and so the coding in the 040 subfield $b$ needed to be "eng".

If you notice individual records coded incorrectly for language of cataloging, feel free to correct them. If you notice patterns of records coded incorrectly, feel free to report those to bibchange@oclc.org. We have tools to use to correct this for large quantities of records at once.
Language of cataloging, which is a required bibliographic element in WorldCat, is also important beyond matching and merging. Controlling links headings in bibliographic records to authority files. Language of cataloging determines which authority file is called on for controlling names and titles. So, you may control names and titles only to the LCNAF when English is your language of cataloging. You can only control to the German Integrated authority file, the GND, when your language of cataloging is German.

Note that only controlling to the LC file is available in Connexion. Within the Record Manager interface, you may search and control to all of the available authority files.

This is about name headings, not subject headings. The specific information for which fields are controllable for each available authority file is documented on the OCLC website on the "Work with Authority Records page". I'll put a link to that in Chat in a few minutes. There is a page for each authority file that details the fields to which controlling may be applied. [https://help.oclc.org/Metadata_Services/WorldShare_Record_Manager/Authority_records/Work_with_authority_records](https://help.oclc.org/Metadata_Services/WorldShare_Record_Manager/Authority_records/Work_with_authority_records)

Regarding Subject headings, one may use subject headings from any schema in any bibliographic record in WorldCat. For example, you are welcome to add Spanish subject headings to a record cataloged in English. Language of cataloging does not apply to subject headings.
Here’s a field that is obscure and rarely used. The 242 field allows the cataloging agency (in other words, your library) to do their own translation of a title and include it in the bibliographic record. This field is optional. It is indexed within WorldCat in the title and keyword indexes. If you have a special collection or particular title in another language in your library, where you believe your patrons would benefit from a translation of the title, feel free to use this field. It contains many of the same subfields as the 245 field --- a, b, c, n, & p. The unique subfield that is distinct is the subfield $y$, where the MARC language code for the language of the translation of the title is included. Only subfields $a$ and $y$ are mandatory in the 242 field. The example shows a title in Spanish in the 245, with an English translation of the title in the 242 field. Note that the cataloger decided not to translate the subfield $b$ – which is perfectly acceptable.

And now back to Shanna.
Thank you, Cynthia. Next, I will cover parallel titles and the ways they are recorded in a bibliographic record.
Parallel Titles (Field 245 Subfield $b)

Not repeatable
Remainder of title, including parallel title(s)
  • In a single subfield $b

Indexed
  • ti: Title Word Index
  • kw: Title Keyword Index
  • ti: Title Phrase Index

Often used in conjunction with field 246

245 10 Café de Costa Rica and its secrets = ¶b Los secretos del café de Costa Rica
¶c Camille Ratton, Yazmin Ross

The 245 subfield $b is the remainder of the title information, which includes parallel titles. These titles are taken directly from the resource in hand.

This subfield is not repeatable, therefore subfield $b should not be repeated when more than one parallel title is given in the field.

It is indexed in the Title Word Index, the Title Keyword Index, and the Title Phrase Index and is often used in conjunction with field 246.

The example here shows a resource in both English and Spanish. An equal sign precedes the subfield $b with the parallel title. An upcoming slide will show the parallel title recorded in field 246.
Parallel Titles (Field 245) continued

**Score:**
- 245 10 Lyrische Suite : ♫ für Streichquartett = Lyric suite : for string quartet =
  Suite lyrique : pour quatuor à cordes / ♫ Alban Berg
- 246 31 Lyric suite : ♫ for string quartet
- 246 31 Suite lyrique : ♫ for quatuor à cordes

**Score:**
- 245 00 Flötenmusik von Komponistinnen : ♫ 13 Stücke für Flöte und Klavier / ♫ herausgegeben von Elisabeth Weinzierl Wächter und Barbara Heller = Flute music by female composers : 13 pieces for flute and piano / edited by Elisabeth Weinzierl Wächter and Barbara Heller
- 246 31 Flute music by female composers


The first example shows a score with titles in German, English, and French, with the English and French equivalents recorded in separate 246 fields.

The following shows a score with parallel titles, other title information, and parallel statements of responsibility. Note in this example the parallel title and statement of responsibility is given after the first statement of responsibility.
Parallel Titles (Field 246: Varying Form of Title)

Repeatable
1st indicator value ‘3’ (no note, added entry)
2nd indicator value ‘1’ (parallel title)
Indexed
- T: Title Word Index
- kW: Title Keyword Index
- Ta: Derived Title Index
- Ti: Title Phrase Index

Often used in conjunction with field 245 $b$

Omit initial articles

245 10 Café de Costa Rica and its secrets → 4b Los secretos del café de Costa Rica / by Camélia Ratton, Yazmin Ross
246 31 Secretos del café de Costa Rica

Field 246 is used to record varying forms of the title, and is often used in conjunction with field 245 subfield $b$.

For parallel titles, the 1st indicator is coded as value 3 (no note, added entry) and the 2nd indicator coded as value 1 (parallel title). Note that CONSER handles parallel titles a little differently in that the 1st indicator is coded ‘1’.

The field is repeatable. It is indexed in the Title Word and Keyword Indexes, Derived Title Index, and Title Phrase Index, and initial articles should not be recorded.

Using the example from the previous slide for the 245 subfield $b$, we see the Spanish parallel title from the 245 subfield $b$ is recorded in the 246 field with the indicator values 3 and 1.
This example shows a resource in multiple languages with each parallel title recorded in a separate 246 field.

As previously mentioned, the 245 $b is not repeatable, so the German and English parallel titles are in a single subfield $b in the 245 field with an equal sign separating each parallel title.

Since there is not a predominate language, the language codes are recorded in English alphabetical order in the 041 field.
Language notes in field 546, but sometimes found in field 500, are used to record language information in textual form that may also be found in coded form in the fixed field *Lang* element and field 041.
Language notes field 546 contains textual information on the language or notation system used to convey the content of the described materials.

The field is repeatable, is indexed in the Note Word Index, and the coded language is contained in the fixed field `Lang` and/or field 041.
This slide shows several examples of the language note in field 546 used in conjunction with the coded language information in the fixed field _Lang_ and field 041.

The first showing a resource in Hungarian with summaries in English, German, and Russian.

The next example is a resource predominately in American Sign Language with minimal English text.

In the last example, the note mentions the original language which is out-of-scope for field 546 and is therefore entered in field 500.
Associated language codes are recorded in field 377.
This field is optional and very rarely used. It contains codes for languages associated with the entity described in the 1xx field of the bibliographic record. This includes the language a person uses when writing for publications, broadcasting, etc., a language a corporate body uses in its communications, a language of a family, or a language in which a work is expressed.

The field and subfield $a$ are repeatable and are indexed in the Entity Attributes Word Index and the Language Word Index.

Please refer to the page for this field in Bibliographic Formats and Standards for guidelines on using this field in relation to fixed field element $Lang$, field 041, and field 546.

The example on the slide shows the author publishes in English and Russian.

And now back over to Cynthia.
Thank you Shanna.

Field 788 is a very new MARC field. It was just added to MARC in 2022 and validated in OCLC in September 2022, so it has been available less than one year. The proposal for the field came from the Library and Archives Canada, since they catalog the same resource in multiple languages of cataloging, and they wanted a way to link those parallel records for the same resource together within MARC. The 788 field is a linking field that allows linking the bibliographic records together. The official definition is noted on the slide. Note that it is repeatable, so if you want to use this field to link to records in many other languages of cataloging you may. And, let me emphasize that it is OPTIONAL. You do not need to use this field unless you find it useful. The indicators control display of notes and display constants for those notes.
If you do plan to use Field 788, here are the required, if applicable, subfields, including the language of cataloging in subfield $e$, the title in subfield $t$, and the record control number in subfield $w$. The Relationship information in subfield $i$, which is optional, is very helpful for humans interpreting what is going on. The examples on the next slide will help to illustrate this.
So, for the reciprocal examples listed here, you'll see that the language of cataloging for the item being cataloged is in the 040 $b, as you would expect. The 788 then links to a description of the same resource, but cataloged in a different language of cataloging, as you will see in the 788 subfield $e. The one at the top is cataloged in English, with the 788 linking to a record with a French description. The one lower on the page is cataloged in French, with the 788 linking to an English language description.

These are examples of how the Library and Archives Canada is using the field. When strictly applying language of cataloging, the subfield $i in these examples would only have text in the same language as that shown in the 040 subfield $b. However, the practice shown on this slide does convey the idea well in both languages.
The last topic to cover today is non-Latin script in bibliographic records. To illustrate the diversity of scripts, I’ve added three examples of different scripts on this slide, all drawn from bibliographic records in WorldCat. Non-Latin script is also referred to as non-Roman script or as vernacular. One may enter, edit, and/or view non-Latin scripts in both the Connexion Client and Record Manager, but not in the Connexion Browser. We’ve received questions sometimes about why a record cannot be edited in the Connexion browser. Often this is because it is a non-Latin script record. The browser does not display the non-Latin fields or allow editing of records with any non-Latin script.
MARC standards, available at the link at the bottom of the slide, provide 2 models for encoding non-Latin script. OCLC uses Model A for bibliographic records. That puts all non-Latin script in 880 fields that may be linked to Latin fields. In cataloging interfaces, OCLC displays these as paired fields, so that you do not see the 880 field coding. The example shows a Georgian script field paired with its transliterated Latin script field. Behind the scenes, the non-Latin field is coded as 880 with a subfield $6 pairing it to the correct Latin field. Likewise, the 100 field contains a subfield $6 pairing it to the correct 880 field. When doing non-Latin script cataloging in Connexion Client or Record Manager, you do not need to enter the subfield $6 coding or the 880 field tag. This happens automatically or perhaps I ought to say automagically, behind the scenes, when you enter non-Latin characters. When the fields are paired, they have a bracket to the left of the field tags, which I've drawn imperfectly on this slide, showing that they are paired in the display. The next slide will show what that pairing really looks like. Pairing and unpairing of fields may be done manually. For records added via batch processes, they are imposed by OCLC programming, which is not perfect in this regard. Check carefully to make sure the machine-imposed pairings are correct when editing records with non-Latin script.
In each of these examples of paired fields, the language of cataloging is English. These paired fields are taken from displays in the Connexion Client. Because the language of cataloging is English, the access points are being controlled by the authorized forms in the LC Name Authority file (LC/NAF). For the first two, the script forms are those found in the LC NAF. The third one is a combination of the LC NAF authorized access point and a free-floating subdivision from LCSH “—Criticism and interpretation.” Note that the subfield $x Criticism and interpretation stays in English when provided in the paired field with the Cyrillic script name.”
These are some reasons catalogers may choose to provide paired fields giving the vernacular script. In all of these examples, the language of cataloging is English, so Latin script is used in the bibliographic records with non-Latin script forms provided in paired fields.

In the first example, the Russian title proper is romanized according to the ALA romanization table, and Cyrillic script is provided in the paired field. You may recognize this as a manifestation of the famous Tolstoy novel War and peace.

In the second example, the authorized access point for the author has been romanized from Greek, and the Greek form is given in the paired field.

In the third example, we have a subject heading for the Chinese politician Chiang, Kai-shek. Those of you who are not familiar with Chinese may not realize this, but his authorized access point is not the ALA romanized form of his Chinese script name. This is a form found in English language reference sources, so it is used according to RDA instructions and LC-PCC policies for names in a non-preferred script. However, the cataloger can still provide a Chinese script form in the paired field, which is helpful to users who know Chinese. The free-floating subdivision—Diaries comes from LCSH and stays in English in the paired field.
OCLC does accept records that contain only non-Latin script, since many libraries around the world do not use Latin script. We are still using Model A, so all the non-Latin fields are actually 880 fields behind the scenes. However, our OCLC system requires a Latin script 245 field in bibliographic records. So, when no Latin script 245 is present, a 245 containing only angle brackets is system supplied to provide that required Latin script field. The first example is, in fact, from a record that is entirely cataloged in the Arabic language in Arabic script from a university in Lebanon.

Looking at the second example, if you have non-Latin script mixed into an otherwise Latin script 245, and you are not supplying a Latin alphabet transliterated 245 field, those angle brackets are system-supplied. This second example shows a mixed field, which is system coded as an 880 field, since it contains Cyrillic characters as part of the field, in this case in subfield $b$. These 245 fields with angle brackets display in Connexion Client but do not display in the Record Manager interface.
The 066 field contains a symbol or a code that tells what non-Latin scripts are present in the record. This field is system supplied, so when cataloging you do not need to look up or enter this data. The subfield $c$ is repeated, as in the fourth example, if there are multiple non-Latin scripts in a record. It can be a very useful clue to tell you what you are seeing in the record. And, the field is indexed, so you can search for the script or use this to qualify a search. For example, if you want to find all the records with Mongolian script held by your library, you could do a command line search in Connexion of: li:XXX where XXX is your OCLC symbol combined with vp:mong. This 066 field only displays in the Connexion Client. It does not currently display in Record Manager. It is still present behind the scenes, and it is searchable. And, it is on the development list for the future for Record Manager to display this field.
Shanna and I both want to thank you in as many languages and scripts as possible for your attention and interest. And now I'll turn it back to Charlene for questions.
On the call today

Charlene Morrison
Senior Data Analyst

Sharon Griffith
Senior Data Analyst

Cynthia Whitacre
Senior Data Analyst

Jay Weiss
Senior Executive Database Specialist

Bryan Bakus
Senior Data Analyst
Thank you!

April Virtual AskQC Office Hours
Data and algorithms and bibs, oh my!

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

Tuesday, 11 April at 9:00 AM Eastern
Thursday, 20 April at 4:00 PM Eastern

Registration and session links available at oc.fc/askqc