

ALA Annual • June 23–28, 2016

# OCLC Authorities Update

**Cynthia M. Whitacre**

Manager, WorldCat Quality



## **Becky Dean, MLS**

Senior Product Analyst, OCLC

---



## **Cynthia M. Whitacre, MLS**

Manager, WorldCat Quality, OCLC

---

# OCLC Controlling: End goal

- All headings in WorldCat are controlled
- The text of the heading is not stored in the bibliographic record, rather, is created when the bibliographic record is retrieved from the database
- The form of the heading is presented is determined based on criteria, such as:
  - User profile in an application will handle the presentation of the heading
    - in a selected language of cataloging
    - based on the language of cataloging of the bibliographic record
- Leverage VIAF, ISNI, and other consolidating identity initiatives to improve the concepts of precision and recall

# Brief history of Controlling

- 2002 First implementation of controlling for the Library of Congress (LC) Names and Subjects Authority file
- 2010-11 Significant improvements made to improve controlling for the LC Names and Subjects Authority file
- 2014 Implementation of controlling using the NTA (Dutch) Authority file
- 2015 Implementation of controlling to the MeSH (National Library of Medicine's Medical Subject headings) Authority file
- 2016 Implementation expected for controlling to the GND (German) Integrated Authority file and to the Māori (New Zealand) Subject Authority file

# OCCLC's definition of controlling

- A bi-directional relationship between a heading in a bibliographic record and an authority record
- This dynamic relationship between a heading in a bibliographic record and a single or multiple authority records enables automatic updates when the established form in an associated authority record changes
- Controlling is a type of linked data

# How this effort differs from VIAF

VIAF: The VIAF® (Virtual International Authority File) combines multiple name authority files into a single OCLC-hosted name authority service.

- *Only includes names and name/titles*
- *VIAF is a compilation presenting all the established headings in a single file*

<https://viaf.org/>

# Goal: Improve discovery for all WorldCat users

- OCLC is working to associate headings from all bibliographic records with authority records and maintain those linked headings
- Language of cataloging determines what file is linked for descriptive headings (e.g., names)
- Subjects are controlled according to tag, indicator values, and/or \$2 code

# Litmus test in adding a new authority file...

- **Requirement: Use the authority file as it is provide to us by the institution/organization**
  - *OCLC does not change the content of the authority records*
- **Understand the contents of the authority records:**
  - so that appropriate rules can be written to associate headings in bibliographic records with authority records using automated functionality
  - to use a set of common indexes to support the future goal of federated searching
  - to be able to process updated authority records and ensure the linked headings in the bibliographic records are handled correctly

# Metadata needed to add a new authority file:

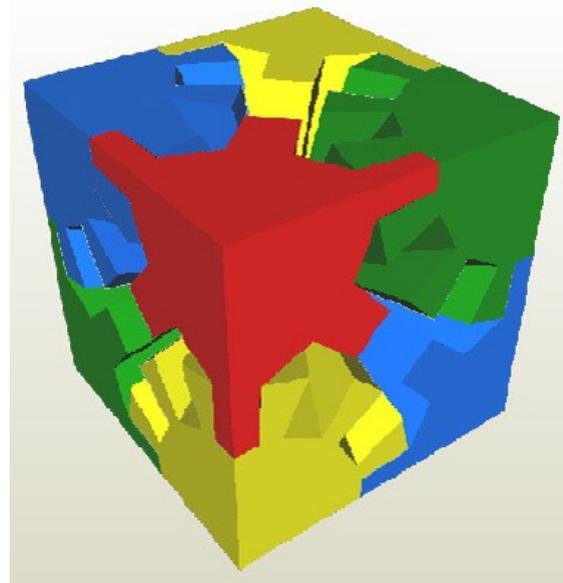
## Requirements for a new authority file:

- Single MARC Organization code for the authority file established and present in the 003 field of each authority record:
  - Library of Congress Name and Subjects Authority File = DLC
  - MeSH (National Library of Medicine) = DNLM
  - NTA Names (Nederlandse Thesaurus van Auteursname) = NL-LeOCL
  - GND (German Integrated Authority file) = DE-588
  - Maori Subject Headings (Nga Upoko Tukutuku) = reo
- MARC-based format (MARCXML or MARC21)
- Appropriate character encoding (MARC-8 or Unicode)

# Analysis of Files

Internal “Fit-for-use” reports allow OCLC staff to review a file and understand its contents.

Often, each file’s contents reveal a rich history based on cataloging practices.



# Key questions: Fit-for-use

Are there data elements that:

- Are incompatible with database requirements (e.g. incorrect length of the 008 field)
- Are incompatible with existing indexing (e.g., presence of fields that are not currently indexed)
- Are used by controlling that use different values than currently defined (e.g., 008/09, 008/32, 008/14-008/16)
- Impact the projected ability for the success of controlling (e.g., \$9 found in 1xx headings)
- Impact the projected ability for updating (e.g., use non-MARC codes in Leader/05)

# Difference in standards and systems...

- Goal is to bring harmony of a multitude of different files into a single catalog
- Understanding that NACO normalization is not appropriate for all files, and embracing diversity of cultures and cataloging history
- Working to understand user needs and workflows—are local practices really needed or are they of the type, ‘we’ve always done it that way...’

# Indexing

- Goal to use a common set of searching and browsing indexes for every file
- Supports a future vision of federated searching across different databases
- Not all indexes are appropriate for all files, but the presentation of the populated indexes is an implementation issue
- Creating indexes that treat diacritics as either significant or insignificant, based on user expectations

## Availability of non-LC authority files is only through WorldShare Record Manager

- Ability to search
- Ability to control headings

These non-Library of Congress linked headings are presented as uncontrolled in Connexion, but contain \$0

# Creation of a unique identifier

- Each authority record must have a unique identifier.
- The identifier uses two pieces of information from the authority record:
  - MARC Organization code found in the 003 field
  - Identifier found in the 001 field
- This unique identifier is expressed in \$0 of a controlled heading
  - NTA Names: \$0(NL-LeOCL)166502634
  - MeSH: \$0(DNLM)D016475

# Examples of *Controlled* descriptive headings in Connexion

- 040 \$b dut
- 100 0 Adele, \$c is Adele Laurie Blue Adkins. \$4 prf \$0(NL-LeOCL)307916901
  
- 040 \$b eng *or \$b is not present (assume eng)*
- 100 0 [Adele](#), \$d [1988-](#) \$4 prf
  
- 040 \$b ger
- 100 0 Adele \$d 1988- \$4 prf \$0(DE-588)135515459

# Examples of descriptive *Controlled* Headings in Record Manager

- 040 \$b dut
- 100 0 [Adele, \\$c is Adele Laurie Blue Adkins.](#) \$4 prf  
*\$0(NL-LeOCL)307916901 presented as mouseover*
- 040 \$b eng
- 100 0 [Adele, \\$d 1988-](#) \$4 prf  
*\$0 is not implemented for Library of Congress authority file*
- 040 \$b ger
- 100 0 [Adele \\$d 1988-](#) \$4 prf  
*\$0(DE-588)135515459 presented as mouseover*

# Examples of *Controlled* subject headings in Connexion

- 040 \$b [any value]
- 650 [blank]7 Mate atewharowharo pukupuku \$2 reo \$0 (Nz)reo2866
- 650 [0,1,2]2 Lung Neoplasms \$0(DNLM)D008175
- 650 [blank]0 [Lungs \\$x Cancer.](#)

*\$0 is not implemented for Library of Congress authority file*

# Examples of *Controlled* subject headings in Record Manager

- 040 \$b [any value]
- 650 [blank]7 [Mate atewharowharo pukupuku](#) \$2 reo  
*\$0(Nz)reo2866 presented as mouseover*
- 650 [0,1,2]2 [Lung Neoplasms](#)  
*\$0(DNLM)D008175 presented as mouseover*
- 650 [blank]0 [Lungs \\$x Cancer.](#)  
*\$0 is not implemented for Library of Congress authority file*

# Challenges...

- NACO normalization is not appropriate for all other files—in more than 50% of the files, diacritics are significant
- Creating a common core set of indexes, that can be used for all files. The goal was to create a set of indexes that could be applied to any authority file and have the data indexed, so that this work could be leveraged in the future for federated searching

# OCLC Controlling: End goal

- All headings in WorldCat are controlled
- The text of the heading is not stored in the bibliographic record, rather, is created when the bibliographic record is retrieved from the database
- The form of the heading is presented is determined based on criteria, such as:
  - User profile in an application will handle the presentation of the heading
    - in a selected language of cataloging
    - based on the language of cataloging of the bibliographic record
- Leverage VIAF, ISNI, and other consolidating identity initiatives to improve the concepts of precision and recall

# Contact Info:

**Becky Dean**

Senior Product Analyst

[deanr@oclc.org](mailto:deanr@oclc.org)

**Cynthia M. Whitacre**

Manager, WorldCat Quality

[whitacrc@oclc.org](mailto:whitacrc@oclc.org)

**Because  
what is  
known must  
be shared.<sup>SM</sup>**