WorldCat Validation release notes, November 2021

Last updated: Fri, 19 Nov 2021 20:51:13 GMT

Release Date: November 19, 2021

Introduction

This release of changes to WorldCat Validation involves the following new features, enhancements, and bug fixes.

• New features and enhancements:
  ◦ OCLC-MARC Validations of New MARC Codes Announced September-October 2021
  ◦ New ISSN Centre Code Validated for Bibliographic 022
  ◦ Bibliographic Field 049 Subfield $a Now Allows Most Latin Characters

• Bug fixes
  ◦ All Authority Validation Rule Sets Now Allow Leader/09 Code “a”
  ◦ Correcting Various Authority Field Relationship Rules
  ◦ Validating Authority Field 682 and Correcting its Relationship with Leader/05
  ◦ Applying the Correct Subfield $4 Relationship Codes to 7XX Authority Heading Linking Entry Fields
  ◦ Unicode Replacement Character (U+FFFD) Made Invalid in WorldCat
  ◦ Using More Sensitive Language in Validation Messages
  ◦ Improved Normalization in Duplicate Field Check

These enhancements and fixes are the result of announcements of new MARC elements and codes by the Library of Congress as well as feedback and requests from members of the OCLC cooperative.

New features and enhancements

OCLC-MARC Validations of New MARC Codes Announced September-October 2021

You may now use all new MARC Codes announced by the Library of Congress in the following LC Technical Notices:

• Technical Notice (September 23, 2021)
• Technical Notice (October 21, 2021)
Gender Code and Term Source Codes
The following source code has been added to the Gender Code and Term Source Codes list for usage in appropriate fields and elements.

- **gsso** - Gender, sex, and sexual orientation (GSSO) ontology (http://www.ontobee.org/ontology/GSSO) [was previously defined for Genre/Form and Term Source Codes list. Usage has been expanded.]

Genre/Form Code and Term Source Codes
The following source codes have been added to the Genre/Form Code and Term Source Codes list for usage in appropriate fields and elements.

- **tpro** - T-PRO Thesaurus der Provenienzbegriffe (Berlin: Staatsbibliothek zu Berlin - Preußischer Kulturbesitz) (https://provenienz.gbv.de/T-PRO_Thesaurus_der_Provenienzbegriffe)

Name and Title Authority Source Codes
The following source codes have been added to the Name and Title Authority Source Codes list for usage in appropriate fields and elements.

- **ntc** - Nederlandse Thesaurus voor Corporaties = Dutch corporations authority file (Den Haag: Koninklijke Bibliotheek)
- **ntd** - Nederlandse Thesaurus voor Drukkers = Dutch printers authority file (Den Haag: Koninklijke Bibliotheek)

Standard Identifier Source Codes
The following source codes have been added to the Standard Identifier Source Codes list for usage in appropriate fields and elements.

- **dma** - Digital musikkarkiv (https://authority.bibsys.no/authority/)
- **orgnr** - Norwegian national organization number (https://www.brreg.no/en/)

Subject Heading and Term Source Codes
The following source codes have been added to the Subject Heading and Term Source Codes list for usage in appropriate fields and elements.

- **nta** - Nederlandse Thesaurus van Auteursnamen (Den Haag: Koninklijke Bibliotheek) (http://data.bibliotheken.nl/doc/dataset/persons)
  Note: This code was previously defined for usage in Name and Title Authority Source Codes. Usage has been expanded.
- **ntc** - Nederlandse Thesaurus voor Corporaties = Dutch corporations authority file (Den Haag: Koninklijke Bibliotheek)
New ISSN Centre Code Validated for Bibliographic 022

One new ISSN Centre Code, 79 for Austria, has now been validated for use in Bibliographic field 022 (International Standard Serial Number) subfield $2.

Bibliographic Field 049 Subfield $a Now Allows Most Latin Characters

Bibliographic Field 049 (Local Holdings) subfield $a (Holding Library) will now allow the input of most Latin alphabetical characters, diacritics, punctuation, superscripts, subscripts, currency symbols, arrows, mathematical operators, and miscellaneous symbols. Field 049 is not retained in the WorldCat Bibliographic record.

Bug fixes

All Authority Validation Rule Sets Now Allow Leader/09 Code “a”

All WorldCat Authority validation rule sets now allow Leader/09 (Character Coding Scheme) to be coded either “blank” for MARC-8 or “a” for Universal Coded Character Set (UCS) (ISO 10646) or Unicode. Previously, value “a” had not been valid in Library of Congress Name Authority and Library of Congress Subject Authority validation rule sets. Please note that there has been no change in Library of Congress Authority record policies, which still allow only MARC-8 characters (https://www.loc.gov/marc/specifications/speccharmarc8.html).

Correcting Various Authority Field Relationship Rules

Several WorldCat Authority relationship rules that had been carried over from Library of Congress Name Authority and Library of Congress Subject Authority validation rule sets were determined not to be appropriate for the general MARC Authority rule set that applies to the Canadiana file and other non-LC authority files, as well as to the future file of FAST authority records. The following relationship rules have been corrected:

- Between the Authority field 008/10 (Descriptive Cataloging Rules) and the presence of Authority field 373 (Associated Group).
- Between the Authority field 008/10 (Descriptive Cataloging Rules) and the presence of Authority field 374 (Occupation).
- Between the Authority fields 151 (Heading – Geographic Name) and/or 181 (Heading – Geographic Subdivision) and the presence of Authority field 034 (Coded Cartographic Mathematical Data).
- Between the Authority field 100 (Heading – Personal Name) First Indicator value “3” (Family Name) and Authority field 008/32 (Undifferentiated Personal Name) value “n” (Not Applicable).

Validating Authority Field 682 and Correcting its Relationship with Leader/05

In all WorldCat Authority validation rule sets, Authority field 682 (Deleted Heading Information) is now valid and its relationship with Authority Leader/05 (Record Status) has been corrected.

Applying the Correct Subfield $4 Relationship Codes to 7XX Authority Heading
Linking Entry Fields

The WorldCat validation rules for subfield $4 in Authority fields 700, 710, 711, 730, 747, 748, 750, 751, 755, 762, 780, 781, 782, 785, and 788 have been corrected to apply the relationship codes from ISO 25964-2: Thesauri and interoperability with other vocabularies — Part 2: Interoperability with other vocabularies, Section 4, Table 1 (EQ, Equivalence; BM, Broader mapping; NM, Narrower mapping; RM, Related mapping) rather than the MARC Code List for Relators.

Unicode Replacement Character (U+FFFD) Made Invalid in WorldCat

In Bibliographic, Authority, and Holdings records in WorldCat, the Unicode Replacement Character U+FFFD (�) now always results in a validation error, eliminating the possibility of input through either online or batch processes. The character was used to replace an incoming character the value of which is unknown or unrepresentable in Unicode. In the new Connexion client 3.0, the character was no longer being displayed or retained, but was instead converted to a blank. This conversion had the unfortunate effect of splitting words and eliminating the visual clue that another character needed to be substituted. Because the character has no valid use in MARC records, OCLC Quality Control staff is deleting all occurrences of the character from WorldCat.

Using More Sensitive Language in Validation Messages

As part of OCLC’s ongoing efforts to focus on equity, diversity, and inclusion throughout all aspects of our work, OCLC has tried to remove the term “master” from documentation and validation messages wherever possible. What used to be referred to as a “master bibliographic record” is now referred to as a “WorldCat bibliographic record.” All validation messages in Record Manager, Connexion browser, and Connexion client now use the more sensitive and inclusive language.

Improved Normalization in Duplicate Field Check

When WorldCat validation checks for duplicate fields in bibliographic and authority records, it has applied the NACO Normalization rules to all relevant fields, the first comma in subfield $a being retained in all circumstances and subsequent commas being regarded as “blank.” Although this makes sense for personal name fields, the results for other fields yields unsatisfactory results. Validation has been changed so that the first comma is retained for normalization only in personal name fields (Bibliographic fields 100, 600, 700, and 800; Authority fields 100, 400, 500, and 700). In all other fields, each comma is converted to “blank” as part of normalization.

Important links

Virtual AskQC office hours

Join OCLC Metadata Quality staff to discuss WorldCat quality issues and cataloging questions. Visit AskQC for information about upcoming office hours, previous office hour recordings, and supporting materials.
Support website(s)

Support information for this product and related products can be found at:

- OCLC Connexion
- WorldShare Record Manager
- WorldShare Collection Manager
- OCLC Community Center
- Cataloging documentation
- Contact OCLC Support