WorldCat Matching release notes, August 2020

Last updated: Thu, 20 Aug 2020 15:37:48 GMT

Release Dates: March through July 2020

Introduction

This announcement of changes to WorldCat Matching software installed between March and July 2020 involves the following improvements, new features, and bug fixes:

- Duplicate Detection and Resolution (DDR) Matching:
  - Rare Materials Exemption for DDR is Now Up to and Including 1829.
  - Improved DDR Matching of Serials with Generic Titles.
  - Improved DDR Matching for Place of Publication.
  - Stricter DDR Matching of Serials When Associated Corporate Bodies have Different Subordinate Units.

- Data Sync/Fingerprint Matching
  - Improved Fingerprint Comparison for Place of Publication.
  - Stricter Fingerprint Matching of Serials When Associated Corporate Bodies have Different Subordinate Units.

- Bug Fixes
  - Fix to the Creation of Fingerprints for Cross-Reference Records.

These improvements have been prompted primarily by feedback from and consultation with members of the OCLC cooperative and were accomplished by the discussion, investigation, and testing work of the matching team at OCLC.

New features and enhancements

Duplicate Detection and Resolution (DDR) matching

Rare Materials Exemption for DDR is Now Up to and Including 1829

In close consultation with the Bibliographic Standards Committee (BSC) of the Rare Books and Manuscripts Section (RBMS) of the Association of College and Research Libraries (ACRL) division of the American Library Association (ALA), the demarcation date for the infrequent automated merging of duplicate records for older materials has been changed. DDR processing now exempts all bibliographic records with dates of production/publication up to and including
1829. Previously, the demarcation had been records with dates of production/publication earlier than 1800. For cartographic materials, DDR continues to exempt records with dates of publication earlier than 1901, as determined in consultation with ALA’s Map and Geospatial Information Round Table (MAGIRT) Cataloging and Classification Committee (CCC). Also still exempted from DDR processing are all records identified in field 040 subfield $e as being cataloged under any of the following MARC Description Convention Source Codes for rare and/or archival materials: amim, amremm, appm, bdrb, cgcrb, cco, dacs, dcgpm, dcrb, dcrmb, dcrmc, dcrmg, dcrmm, dcrmmss, dcrms, dmbsb, enol, estc, gihc, iosr, ohcm, rad, ma, vd16, vd17.

**Improved DDR Matching of Serials with Generic Titles**

DDR matching for generic serial titles now treats the presence of an “author” in 1XX or 7XX in one record and the absence of such an “author” in another record as a mismatch. Previously, matching was occasionally allowing serials with identical generic titles but emanating from different corporate bodies to match.

**Improved DDR Matching for Place of Publication**

In conjunction with the improvements to fingerprint matching reported below under the heading “Improved Fingerprint Comparison for Place of Publication,” DDR place of publication matching was also fine-tuned to better differentiate identically named places in different states or other jurisdictions. Previous to these changes, DDR was occasionally incorrectly matching such resources as yearbooks with the same title from similarly named high schools in different states.

**Stricter DDR Matching of Serials When Associated Corporate Bodies have Different Subordinate Units**

When matching serial records, DDR is now stricter in comparing the names of corporate bodies, mismatching when subordinate units have any difference. Previous to this change, DDR was occasionally allowing a match in cases where associated corporate names had only slight differences.

**Data sync/Fingerprint matching**

**Improved Fingerprint Comparison for Place of Publication**

Fingerprint matching now takes the place of publication into more careful and detailed consideration by incorporating most of the relevant comparisons that have long been used in DDR and had been used previously in pre-Data Sync batchloading processes. Summarizing briefly, these include the extraction, normalization, and comparison of the first 260/264 subfield $a. If the place of publication is missing or mismatching, the codes in 008/15-17 are then compared. Prior to these improvements, fingerprint matching was performing a much more rudimentary comparison of place of publication.

**Stricter Fingerprint Matching of Serials When Associated Corporate Bodies have Different Subordinate Units**

When matching serial records, fingerprint is now stricter in comparing the names of corporate bodies, mismatching when
subordinate units have any difference. Previous to this change, fingerprint was occasionally allowing a match in cases where associated corporate names had only slight differences.

Bug fixes

Fix to the Creation of Fingerprints for Cross-Reference Records

When a WorldCat record has been either merged into another or deleted outright, the fingerprint entry that had been created for that record remains in the internal table of fingerprints. This allows a cross-referenced (merged) or deleted WorldCat record to be included in a list of candidate matches. Until this fix, it was possible for an incoming Data Sync record to match a WorldCat record without any confirmation. Now, both a cross-referenced (merged) OCLC number and a deleted OCLC number are always considered to be a mismatch to any incoming record because the underlying data that had generated the fingerprint is either fundamentally altered by the merge or absent because the record has been deleted. In WorldCat displays, the OCLC numbers of cross-referenced (merged) records appear in field 019 of the retained record. The OCLC numbers of records that are deleted outright are not retained in WorldCat.

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.