Introduction

This release of changes to WorldCat Matching involves the following improvements to Duplicate Detection and Resolution (DDR):

- Improved DDR Comparison of Medium of Performance Differences for Musical Scores.
- Improved DDR Matching of Extent in Field 300.
- Improved DDR Matching of Online Serials.
- Improved DDR Matching of Roman and Arabic Numerals in Titles.

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

New features and enhancements

Duplicate Detection and Resolution (DDR) matching

Improved DDR comparison of medium of performance

Matching software can now better distinguish between musical scores representing different instrumentations. Since it was expanded to include bibliographic formats beyond books a decade ago, Duplicate Detection and Resolution (DDR) software has taken medium of performance into consideration for scores. Before the recently installed correction, the software was accepting a match on field 028 as enough evidence that the medium of performance was also a match, on the assumption that publishers generally used different plate and/or publisher numbers to distinguish among different instrumentations. When it was discovered that some publishers assigned the same identifiers to scores for different instrumentations, the algorithm was improved to explicitly consider the medium of performance even in cases where 028 fields match.
Improved DDR matching of extent in field 300

DDR matching can now differentiate better between single volume extents and open entry or multivolume extents by recognition of additional variant presentations of the data in field 300. Previously, DDR was occasionally incorrectly equating certain ways of expressing the extent of a single volume resource with certain ways of expressing the extent of a multivolume resource.

Improved DDR matching of online serials

DDR matching now identifies and merges more duplicate records for online manifestations of continuing resources. Previously, such records with identical OCLC symbols in field 040 subfield $c were purposely not being merged under the supposition that the institution intended separate records for legitimate reasons. In the time since the 2009 implementation of this rule, multiple improvements especially to title and “author” matching have proven themselves to be more effective in making accurate matches to such an extent that the old rule has become counterproductive. Removal of this outdated rule means fewer duplicates for electronic serials.

Improved DDR matching of Roman and Arabic numerals in titles

DDR matching now both better differentiates and correctly matches Roman numerals in title field 245. DDR also now more accurately recognizes when Arabic numerals and Roman numerals represent equivalent numeric values. Previously, under certain specific circumstances involving Roman numerals, DDR was occasionally incorrectly matching some different numeric values and overlooking some equivalent numeric values.

Data sync/Fingerprint matching

There are no data sync/fingerprint matching updates at this time.

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.