Matching to records in WorldCat


OCLC uses specific methods depending on collection type to match data to records in WorldCat.

Matching for knowledge base collections

Titles in the WorldCat knowledge base are matched to WorldCat records based on their associated metadata including:

- the format, title of the item, standard identifiers (like ISSN/ISBN), publisher, and author (for ebooks)

The matching process first identifies all records in WorldCat that match the provided metadata and then narrows the match to a single record based on three criteria or rules:

1. Select a matching record with the same format as the item.
2. When there are multiple records that match the item profile, select the record with more holdings.
3. When the prior two criteria are the same, select the record with the smallest OCLC number.

Alternative processes

Certain collections have been matched through an alternative process. Such collections correspond to content providers who are participating in one of OCLC's cataloging services. OCLC number matches for these collections are either supplied by the content provider or merged into the WorldCat knowledge base from sources internal to OCLC. As a result, they may not follow the exact same matching process described above.

If title data does not include an OCLC number

If title data does not include an OCLC number and the knowledge base cannot add an OCLC number using its matching process, a record will not be output for that title. This can include cases where:

- No records exist for the item in WorldCat
- Records exist but are not classified as electronic resources in a way that corresponds to the knowledge base’s matching process because of the complexities of cataloging electronic items

OCLC number coverage

If a matching OCLC number becomes available at a later date, the title will be matched and its record will be delivered automatically.

More information

For more information about matching to records in WorldCat and about requirements for full-text link display, see:

- Display full-text links
- MARC Record delivery troubleshooting
Matching for cataloging partner collections

Matching your electronic order information with your collection in Collection Manager

Collection Manager uses the account number values and collection ID values to match the collection you create in Collection Manager with the electronic order information your material provider sends. Your cataloging partner collection is then matched to records in WorldCat.

Matching your holdings with records in WorldCat

Cataloging partner collection matching emphasizes matching a unique number in the incoming material provider record with a unique number in the WorldCat record. Of the unique numbers used, cataloging partner collection matching algorithms prefer OCLC numbers when available. Other unique numbers include LCCN, ISBN, ISSN, etc. In order to ensure better matches, cataloging partner collection matching combines these unique numbers with bibliographic information such as title, publication dates, material types, and language of cataloging to distinguish among similar but distinct manifestations.

Since the emphasis is on unique numbers, cataloging partner collection matching rarely confirms more than one WorldCat record as a match to the incoming material provider record. When more than one record is confirmed, cataloging partner collection uses record resolution algorithms to select the best record. The record resolution algorithms select records based on the highest encoding level, the cataloging source, and records that have been reviewed by national cataloging authorities based on the 042 Authentication Codes.

Matching for data sync collections

Note: This section is an overview of the data sync matching process. It does not cover other OCLC matching processes (e.g. ILL, knowledge base, Duplicate Detection and Resolution (DDR), etc.). It does not include all of the processing details, anomalies, and exceptions that may occur during fingerprint matching.

Inaccurate or incomplete cataloging and coding, seemingly small differences between records, and certain types of inconsistencies within a record may cause machine algorithms to act in unexpected ways. No automated matching algorithm, regardless of its sophistication, can consider and correctly interpret all of the subtleties built into a bibliographic record with the accuracy that a human cataloger can.

Introduction

After bibliographic records are created, sent to OCLC, and pre-processed, they are sent to matching. Matching provides an automated way to match incoming records to a single WorldCat record so that library collections can be made part of WorldCat for efficient and effective cataloging, resource sharing, library management, research and Discovery.

There are two basic steps for matching bibliographic records:

1. **Retrieving candidate records from WorldCat.** The matching code uses pattern recognition—in the form of
“Fingerprints”—to retrieve WorldCat records. Data in incoming records are composed into Fingerprints. Those Fingerprints are matched with Fingerprints composed from WorldCat records, and potential candidates are retrieved based on matching Fingerprints. The OCLC control number serves as the fingerprint element in the first attempt to retrieve a record. While the matching process has always used this setting, it now plays a larger role in identifying the best record.

2. Comparing records. Next, matching determines if any of the candidate records represent the same manifestation of an item as the incoming record. It does this by comparing data in the records. If the record is retrieved using the OCLC control number, title and material type are compared. If this finds a match, the record moves on to processing. If this does not find a match, other fingerprint elements are used to find a match. If multiple records meet the matching criteria, the records are sent to Resolution. In Resolution, additional criteria are used to select the best WorldCat record to use as a match.

Data elements used in matching

Data elements can be single MARC fields and subfields or combinations of MARC fields and subfields. The data elements are used in both retrieving records and comparing data. Algorithms determine how data elements are constructed for retrieval and how they are used in comparing records. Matching emphasizes the primary data elements. Only in certain limited circumstances are 5XX notes used in matching.

Data elements used as the primary source of retrieval and comparison for matching include, but are not limited to, the following:

- Date of publication
- Language of cataloging
- Title

[Table of MARC fields and subfields used in matching]
- "Unique" Numbers including OCLC Numbers, ISBN, ISSN, etc.
- Physical Material Type
- Dates of Publication
- Language of Cataloging
- Title
- Author
- Edition
- Publisher
- Extent

In addition, other data elements used for specific situations and formats include, but are not limited to, the following:

- Archival designations
- Serials designations
- Places of Publication
- Photocopies
- Size
- Musical parts
- Recording dates for audio and moving image recordings
- Performers, directors, conductors, etc., particularly for audio and moving image recordings.

**Record resolution**

When matching finds more than one WorldCat record that could represent the manifestation of the incoming record, it sends the records to Resolution. Record resolution is a separate process that attempts to choose from among the incoming record and the WorldCat records the best candidate for record retention based on algorithms that include but are not limited to the source of the record (field 040 subfield $c), certain authentication codes in field 042, Encoding Level, and the number of holdings.

**Final actions**

After Resolution selects the best record, data sync sends the record to validation and on to Final actions. Final actions include adding a WorldCat record, adding a holding, canceling a holding, replacing a record, and/or transferring additional data not already present in the WorldCat record.