

Matching/Deduplication Criteria

The current program proceeds sequentially through the following 5 steps and in this order:

OCLC Records:

1. By OCLC Control Number [mandatory] only. Format Indicator is ignored.

Note: All OCLC Control numbers shall have the “ocm” prefix otherwise the record will be processed as a non-OCLC record. Exception: Because Innovative Interfaces (III) systems do not retain the “ocm” prefix, these records will be evaluated using the following criteria:

- Control number shall be in the correct field (according to the profile), and
- At least one of the following shall be true:
 - Presence of “OCoLC” in Field 003
 - Presence of 049
 - Presence of 019

Non-OCLC Records:

2. LCCN, Format Indicator [mandatory], 1st 10 characters of Title (245a)
3. ISBN, Format Indicator [mandatory], plus 1st 10 characters of Title (245a)
4. ISSN, Format Indicator [mandatory], plus 1st 10 characters of Title (245a)
5. If matching by number with text validation fails, matching would be via TEXT on:

Full Title: 245 \$a [mandatory], \$b, 1st 7 characters of subfield \$h, 1st 20 characters of subfield \$p

Physical Description [mandatory], (Field 007 Positions 00 & 01)

Author (1st 80 characters)

Language (Field 008 Positions 35-37)

Form of item (Field 008 Position 23)

Publication Date: 260 \$c 1st 5 dates (ignores brackets and alpha characters)

Publisher Name: 260 \$b 1st 5 words (ignores initials)

Format Indicator [mandatory]

Note: the text validation criteria for items 2-5 above can be changed. Initial articles are ignored when matching on title. All opportunities for matching by control number (steps 1 - 4) are exhausted before attempting to match by text alone (5). If title match fails, the record is added as a new record to the Catalog.

Mandatory and Non-mandatory Criteria

All mandatory items shall be present and match. All nonmandatory items shall be present otherwise a match is precluded.

Format Indicator - Leader Positions 6 and 7

The absence of a format indicator in only one of the records precludes a match, but its absence in both records does not preclude a match.

iCONN.org – Connecticut’s re-search engine

Request for Bid

Matching Subfields for Number Matches

<u>Input record subfield</u>	<u>Master record subfield</u>	ACTION
\$a	\$a	match occurs
\$a	\$z	match occurs - only holdings are updated
\$z	\$a	match occurs
\$z	\$z	no action - this condition is not checked for

PUBLISHER VALIDATION (260 \$b)

If 260 \$b is requested as a validation point, a two-way keyword is used for the first 5 keywords in the subfield (initials are ignored).

The match looks for the first significant keyword in the incoming record's 260 \$b subfield, and matches it to that keyword anywhere in the potential match record.

If a match fails, the same process is reversed and the match is tried again. Multiple instances of \$b are searched for a match.

For instance, a match like:

260 \$a New York : \$b A. E. Knopf, \$c 1988.
260 \$a New York : \$b Knopf, \$c 1988.

would occur because initials are ignored in selecting the keyword to be used.

In another example, the keyword match is two-way. Two publishers will be found to be a match:

260 \$a New York : \$b Beginner Books: \$b distributed by Random House, \$c 1997
260 \$a New York : \$b Random House, \$c 1997.

When the matching is reversed: The first pass looks for a match using the keyword BEGINNER and fail, but when the process is reversed, the match is found using the keyword RANDOM, then finding the match in the second \$b of the matched record.

PUBLICATION DATE VALIDATION (260 \$c)

Validates for the first (x) dates in the 260 \$c subfield. Brackets and alpha characters are ignored. The match is a two-way. For example, the following 260 \$c samples would be seen as a match if the processor chooses to validate for four dates in the 260 \$c

260 \$a New York : \$b Knopf, \$c 1988, c1982
260 \$a New York : \$b Knopf, \$c 1982.