

5. How Batchload Processes Records

Last revised July 2011

- 5.1. Processing.....2
- 5.2. Preprocessing.....8

5.1. Processing

Please note: The descriptions of processing in this section are **default settings** for batchload projects. Some projects, specifically those that database specialists review manually before processing, may have other settings, depending on the results of evaluation by the OCLC database specialist assigned to the project and on the specialist's discussions with the library.

See a more detailed account of how records are matched in WorldCat for processing in appendix A, "Matching Records for Batchload."

5.1.1 Unique key matching

OCLC number. Properly formatted OCLC control numbers are mapped to field 035 \$a and prefixed with (*OCoLC*) for use by batchload software.

When you order a batchload project that uses MARC records, you are asked to select which, if any, of the following qualifiers to apply to OCLC number matching:

- Derived title (fields 245, 246, 247) (**default**)
- Date (008/07-10)
- Material type (Leader/06)
- Language of cataloging (field 040 \$b)

Qualifiers cross-check other data in matched records. See more about whether/how to [select OCLC number qualifiers](#).

Other unique keys. The following is a list of unique keys used for matching. The OCLC number qualifiers listed above are applied.

- 010 \$a (LCCN)
- 016 \$a (National bibliographic control number)
- 020 \$a (ISBN)
- 022 \$a (ISSN)
- 024 \$a (Other standard number)
- 027 \$a; \$z and 088 \$a; \$z (Report number)
- 028 \$a and 262 \$c (Publisher number)
- 029 \$a (Other system control number)
- 030 (CODEN)
- 856 \$u (Uniform resource identifier)

5.1.2 Extended matching

When an OCLC control number is not available, or when the qualifiers rule out the OCLC number as a match, or when a single match is not retrieved by the other unique keys, the system uses extended matching based on additional data in the body of the record (author, title, publisher, and more).

5.1.3 Enriching WorldCat automatically

Batch processing automatically merges some data in records if the data enhances the records (e.g., fields 029, 040, 938, etc). LC call numbers and subject headings, Dewey call numbers, content notes, summaries, or non-English subject headings in your records may be merged into matching WorldCat records that lack them. Other data may be merged at the discretion of the database specialist assigned to your project.

Note: For matched records, an algorithm can be used to determine whether the library's record or the existing WorldCat record is retained as the master record. An algorithm and processing specifications determine which fields of the records are merged to create the final master record. See appendix B, "Merging Specifications" for more information.

5.1.4 Adding unmatched records

When you order a batchload project and your records are MARC, you can select the option to add your unmatched records to WorldCat as original cataloging.

If you request that your unmatched records be added, the records must meet these standards and criteria to be added:

- National standards for cataloging, including those for descriptive cataloging and forms of headings.
- OCLC K-level standards for completeness, as described in *Bibliographic Formats and Standards*:
<http://www.oclc.org/bibformats/en/onlinecataloging/default.shtm#BCGGBAFC>
- Titles that can be used for matching. If field 245 is missing or lacks both \$a and \$k, the record cannot be added to WorldCat. Such records go into the unresolved file. (See chapter 6 "Exception Data (Records Remaining)" for more information.)

See more details about [sparse records](#) below.

- Batchload validation errors that are at an "acceptable" severity level (determined by the system) or that can be fixed for processing using preprocessing routines.

See more about [invalid records](#) below.

- Records from vendors who have either given your library or OCLC permission to redistribute the records.

If the vendor has not given redistribution permission, these records are considered to be "restricted use" records and cannot be added to WorldCat. See 3.2 "Guidelines for using the order form" for instructions about how to treat these records for batchloading.

More about records that are too “sparse” to be added to WorldCat

Definition. Sparse records are bibliographic records that lack sufficient data to identify a **manifestation** of an item. For example, a record for the book *Dombey and Son* by Charles Dickens that has only the title and author’s name does not identify the precise manifestation of the book by publisher, date of publication, language, format (for example, Braille or large print), editor, and other data that would distinguish the particular manifestation of *Dombey and Son*.

Used for processing? Sparse records that OCLC receives for a project are not added to WorldCat or used to replace WorldCat records. They are used if they match in WorldCat to set or cancel holdings. Unmatched sparse records are placed in the unresolved records file.

Conditions that define a record as too sparse

- a. Field 245 lacks both subfields \$a and \$k
- b. Contains field 245 with subfields \$a or \$k but does not have any of the following fields:
 - o 250 – 490
 - o 773
 - o 020
 - o 022
 - o 024
 - o 028

Exception: Even if a record has the 024 and a field 245 with subfield \$a or \$k, if the first indicator value in field 024 is **8** and the record has no other fields, it is considered to be sparse.)

With the exception noted above and with the two exceptions below, if any of the fields in this list are present along with field 245 \$a or \$k in a record, it is a candidate (other criteria must be met) for adding to WorldCat.

- c. Field 245 \$a or \$k contains the following text (case-insensitive):
 - o *dummy title*
 - o *no title found*
 - o *no title*
- d. Data in field 245 \$a or \$k does not contain alphanumeric characters. Examples of invalid titles:
 - o 245 00 ... / \$c Henry Thomas.
 - o 245 10 .
 - o 245 00 *** / \$c par M. ***
 - o 245 10 ##
 - o 245 00 ???
 - o 245 00 \$c Sotheby’s.

Exception: A record that contains field 245 with only angle brackets in subfields \$a and \$6 is not considered to be sparse. It is a candidate (other criteria must be met) for adding to WorldCat.

Example of acceptable title with angle brackets:
245 00 \$6 880-01 \$a <>

Note on sparse records for users of SCIPIO

A SCIPIO record is too sparse to add under the following condition **in addition to** conditions **a**, **c**, and **d** listed above:

- e. Lacks fields 024, 033, and 110.

More about invalid records

Batchload Error Validation software checks MARC 21 content designation and record format (structure) in records at several points during processing, using a less rigorous subset of the online validation used in Connexion.

Invalid records can contain invalid codes, fields, indicators, and/or subfields, lack required fields, or repeat non-repeatable fields.

Validation includes categorizing errors according to severity and may identify errors that can be fixed. System rules for identifying the severity level of validation errors determine whether records with errors can be used for matching in WorldCat or can become candidates (other criteria must be met) for adding to WorldCat if unmatched.

See details in appendix C “Batchload Validation for Bibliographic Records.”

5.1.5 Setting and canceling holdings

Caution: When your library’s holdings are deleted from OCLC master bibliographic records, any attached local records are also deleted, including local holdings records (LHRs) and local bibliographic data records.

Batch processing uses record status to determine whether to cancel or set your holdings in WorldCat. (**Record status** is labeled “[Rec stat](#)” in the fixed field display in bibliographic records; it is also referred to as Leader byte 05 or 008/05.)

When record status contains the value **d**, your institution’s holding symbol is canceled (deleted) on matching WorldCat records.

For any value other than **d**, your holding symbol is set (added) on matching and newly added records.

Depending on which one of the following types of holdings actions you select when you order a batchload project, this is how holdings are processed:

- **Cancel for all.** Batch preprocessing forces the record status value **d** for all records submitted, regardless of the value in your records, so that your holdings are canceled on all matched WorldCat records.
- **Set for all.** Batch preprocessing changes any occurrence of **d** in record status to **c**, so that your holdings are set for all matching or added records.
- **Set or cancel holdings based on the record status value.** Batch preprocessing accepts the record status value in your records, so that your holdings are canceled for matched records with value **d**, and holdings are set for matched records with any other value.

Note: For this third type of holding action only, preprocessing checks record by record. For the other two holding actions—cancel for all or set for all—values are forced for all records.

When record status value cannot be used. If your library's local system does not have Leader Byte 05 for record status implemented, you cannot use the value **d** to indicate cancel holdings. Instead, you can order two separate projects and send two separate sets of records, one project to cancel holdings on all records sent and another project to set holdings on all records sent.

5.1.6 Replacing records in WorldCat (one-to-one)

Under special circumstances, OCLC can replace a WorldCat record that your library contributed with a new record from your library, if the record has not been enhanced by other libraries. This type of processing may be used for bibliographic records that represent unique items held only by a single institution, such as archives and special collections.

You can request one-to-one replace in the batchload order in the free-text box for the question "Do you have more information to add about your project?" After you place the order and you receive your project ID ("P" number, for example "P012345"), a database specialist will contact you to work out the details.

Note: Regardless of the value in 040 \$c, when you send a record, any record that gets added has 040 \$c set to the OCLC institution symbol of the contributing library. (Originally, the value might have been the MARC institution code of the contributing library, or the OCLC symbol/MARC code of the institution that originally transcribed the record into machine-readable form.)

Detailed rules for replacing records

When your records are processed, for those that match, the WorldCat records are replaced according to the following rules:

- **Cataloging source rule:** The symbols in field 040 (Cataloging Source) \$c (Transcribing Agency) of the library's record and the matching WorldCat record must match. If they match **and** the following modification rules are met, the record is replaced. If they do not match, the WorldCat record is not replaced. (040 \$c is the OCLC symbol of the organization that created the record in WorldCat.)
- **Modification rules:** If the WorldCat record was modified, it contains 040 (Cataloging Source) \$d (Modifying Agency). To be replaced, records need to have the symbol types listed below. If 040 \$d has any symbols other than those described below, the WorldCat record is **not replaced**. This rule protects modifications made to the record by all other OCLC members.

Only records with the following kinds of symbols in 040 \$d **are replaced**:

- The symbol in 040 \$d matches the symbol in 040 \$c. This rule assumes that the modifying agency is the library that first created the record and therefore is allowed to replace it.
- The symbol in 040 \$d is an OCLC quality control processing symbol (OCL, OCLCA, OCLCG, OCLNG and OCLCQ). This rule assumes that the record creator's information takes precedence over quality assurance corrections.
- The symbols in 040 \$d are those of vendors (listed below) for whom 938 fields are being added. This rule assumes that the record creator's information takes precedence over vendor-added information.

ACTRA	BTCTA	DKAGE	I5B	RECBX
ALBRS	BWK	DON	LEILA	SBACH
AUXAM	BWX	DVDSN	LNDMK	TRCLS
BAKER	C3L	ERASA	NEIDC	TZT
BBCAA	CASSC	ERASP	OHX	WEINB
BDX	CDX	EVIEW	PL#	YDXCP
BOLIN	CNPIT	GGVRL	PLAYA	
BLICP	DG1	H7K	QUICK	

Note: Records that are considered to be too sparse are not used to replace records. See more about sparse records above.

5.1.7 Storing local bibliographic data

When you order a bibliographic batchload project to support full-subscription WorldCat Local or to support migration to Web-scale Management Services (WMS), you can request that OCLC extract local bibliographic data from your records for display to users via your OCLC discovery interface.

If you select this option, batchload creates local bibliographic data records attached to the associated WorldCat bibliographic records. The local bibliographic data is visible only through WorldCat Local or WMS and can be maintained through batchload. You cannot view or edit it using Connexion or any other interface.

Results of extracting local bibliographic data are reported with the other results of your batchload project in File Processing Summary Reports for each file processed and in your final Project Processing Summary Report when your project is complete.

Note: If you later delete holdings on WorldCat records that have your library's local bibliographic data records attached, the local bibliographic data is also deleted.

See more information in [Storing Local Bibliographic Data via Batchload for WorldCat Local or WMS](#).

5.2. Preprocessing

Batchload preprocessing helps to prepare records and resolve potential problems with your library’s data that may prevent successful processing. If a batchload specialist needs to review your project, s/he may analyze a sample of the data, run tests on your records, and then write and run preprocessing routines on the records. The routines transform data in ways that are needed for matching records, setting and canceling holdings, resolving records, merging data, and updating WorldCat records. Data in the records may be moved, normalized (that is, made consistent), or reformatted.

Caution: Preprocessing is a "best efforts" attempt to prepare your data for processing. Batch Services cannot guarantee that these preprocessing routines (1) will lead to successful processing or (2) will not adversely affect the quality of coding in your records.

This preprocessing section gives examples of errors and how records may be preprocessed in the attempt to fix them. You can use information in this section to evaluate your library’s records for potential problems **before** you order a batchload project. The best case, of course, is for OCLC to receive records that are free of these problems, whenever possible.

See part 6 “Exception Data (Records Remaining)” in this Guide for more about records remaining that have been processed, usually multiple times using various routines, but are returned to you because they could not be matched or added in WorldCat.

5.2.1 Critical errors and preprocessing routines

The following table describes “bad,” missing, invalid, or un-matched elements in records that cause critical errors. These errors result in records that are returned to you as “exception data” (defined above). The table also lists preprocessing routines that are used to try to fix or mitigate the errors.

Critical Error	Description of error	Preprocessing routine does this...
“Bad” record leader and record directory	The leader or directory may have incorrect data in it, may be too long or short, or may contain illegal characters.	Re-blocks and reformats the records.
001 field missing (See 5.2.2 below for more details.)	Batchload software uses the unique local system control number in field 001 for de-duplication and post-processing reports.	Supplies the 001 field and the local system control number from another field.
Multiple 001 fields (See 5.2.2 below for more details.)	Batchload software uses the unique local system control number in field 001 to de-duplicate your files.	Removes all but one 001.
008 field missing	The 008 field is the fixed field of the MARC record.	Codes the 008 based on the content of variable fields in the record. The quality of the coding varies.

Critical Error	Description of error	Preprocessing routine does this...
008 incorrect length	Often, if the format of the record is incorrect, the fixed field is too long or too short.	Codes the 008 based on the content of variable fields in the record. The quality of the coding varies.
Date1 missing from 008 field or invalid	Date 1 is required and must be valid. Invalid data in Date1 includes, for example, dashes, question marks, spacing underscores, alpha characters, blanks for unknown dates.	Uses dates in the record (for example, 260 Ꞥc) or supplies a dummy date. The outcome varies.
245 field missing	No title	May provide a dummy title, but most often does not attempt to correct missing titles.
Invalid Rec Stat (See 5.2.3 below for more details.)	Batch processing uses Record Status to identify records for canceling holdings.	Forces Rec Stat to value d for cancel holdings projects.
Invalid Type (Rec Type)	Type and Bib Level combined determine the format of the record, which must be valid.	May try to correct the coding based on the content in variable fields. The quality of the coding varies.
Invalid Bib Level	Same as for invalid Type.	Same as for invalid Type.
Invalid escape sequences	These errors are usually caused by character encoding problems. For example, individual fields may contain Microsoft encoding instead of UTF-8 Unicode or MARC-8.	Can only rarely correct these errors.
Contains 066 field but no 880 field(s)	Field 066 identifies non-Latin scripts. 880 fields contain the non-Latin script data. Special processing is required for non-Latin scripts.	Can remove the 066 field but cannot supply missing text in 880 fields.
Contains 880 field(s) but no 066 field	Same as for the error directly above.	Same as for the error directly above.

5.2.2 More about the 001 field

When you order a batchload project, you enter the field in your records that contains the library's local system control number. If needed, as part of your project setup, OCLC moves the number to field 001, so that the number can be used to:

- Accurately de-duplicate your files.
- Produce Cross Reference (XREF) reports that list each OCLC control number across from the correct corresponding local system control number to help you index the OCLC numbers in your local system. Indexing the OCLC control number links your library's records to WorldCat.

In the MARC 21 standard, field 001 (Control number) contains the control number assigned by the organization that creates, uses, or distributes the record. However, sometimes libraries do not use field 001 for their own control number. For example:

- A library may retain whatever data is in the 001 from source records (for example, LCCNs, OCLC control numbers, or vendor numbers).

Or

- Field 001 may be missing from a library's records—the library may delete the field, or the field is not used in the local system, or it is excluded from records exported from the local system.

Before your records are processed by the batchload matching software, preprocessing routines can create an 001 field, if missing, and can supply the unique local system control number (sometimes called the bibliographic record ID) from the field you identified in your order form (for example, from a 9xx local field).

Example 1

Batch Services received a file in which 19,877 records were missing field 001. Batch preprocessing supplied the 001 field and inserted the local system control number from field 904.

Other records in this file contained 001 fields, but the database specialist assigned to the project tested the file and found the same number in the 001 for different titles: the numbers were not unique.

For example, four records had the same number in field 001, but the titles showed that they are not duplicate records:

001 field	Title (245 field)
bwi07988692	TaeBo live!=bAdvanced Workout_h[videorecording]
bwi07988692	TaeBo live!=bBasic Workout_h[videorecording]
bwi07988692	TaeBo live!=bInstructional_h[videorecording]
bwi07988692	TaeBo live!=b8-Minute Workout_h[videorecording]

To correct the misleading duplicate numbers, the database specialist ran a preprocessing routine that replaced the data in the 001 of each record with the unique local system control number from another field. When batch processing started, the software treated these records correctly as non-duplicates and processed them.

5.2.3 More about invalid record status

The value **d** in record status (Rec Stat) in a record—Leader Byte 05 (008/05)— indicates that institution-level holdings should be canceled on that record.

If your local system can output a file with record status **d** for those records on which you want to cancel holdings, you can send a single MARC file, and batch processing can use the record status value to set holdings for matching records with any value except **d** and cancel holdings on matching records with value **d**.

If your local system cannot output records with a Record Status **d**, you can order separate projects for two separate sets of records:

- One project for canceling holdings on all records sent that match and one for setting holdings on all records sent that match. See more about how preprocessing prepares these records in section 5.1.5 above.

Or

- Two separate numeric search key projects, one listing search keys for canceling holdings on matched records and a separate list of search keys for setting holdings on matched records. See more about numeric search key projects in 2.2 “Project types.”

5.2.4 Other common errors

The following table describes other errors, how they can be prevented, and whether they can be addressed by preprocessing routines.

Type of common error	Description of error	Prevention	Preprocessing routine available?
Tag order	Fields 001, 003, 005, and 035 are located at the end of records.	These fields should be in tag order.	No
UTF-8 Unicode character encoding	Records are in UTF-8 Unicode character encoding but do not contain the proper designation in Leader byte 09 (008/09).	UTF-8 Unicode character encoding must be designated by a in Leader 09. If Leader 09 is blank, encoding indicates MARC-8.	No
Country codes (008/15-17)	Three-character country codes are used in Leader bytes 15 to 17 (008/15-17) instead of the two-character code. And/or country codes include uppercase alphabetical characters.	Leader bytes 15 to 17 should use two-character codes. Codes should all be in lowercase.	Yes
Language codes (008/35-37)	Language codes in Leader bytes 35 to 37 (008/35-37) have uppercase alphabetical characters.	All alphabetical characters should be lowercase.	Yes
Duplicate records (does not apply to group projects)	Preprocessing found duplicate records in files, identified by the local system control number in the 001 field. No other checks are made to verify that the records are true duplicates.	Database specialists must manually examine the duplicates. If found to be true duplicate records, the latest version of the record will be processed.	No

Type of common error	Description of error	Prevention	Preprocessing routine available?
040 ‡b	OCLC encourages its membership to include the language of cataloging (040 ‡b) in records. If not included, the language of cataloging is English by default. Therefore, the lack of a code in a record with cataloging data in a non-English language misrepresents the cataloging.	Include a language code in 040 ‡b if any cataloging data is in a language other than English.	Most often, no (Preprocessing cannot supply a language code in 040 ‡b unless it applies to all records.)
040 ‡e	Cataloging description MARC code was noted in the 040 ‡e. ppiak = Verona, Eva. Pravilnik i prirucnik za izradbu abecednih kataloga (Zagreb: Hrvatsko bibliotekarsko drustvo)	The rules for descriptive cataloging used to create the records were other than AACR2, AACR1, ALA 1908, ALA 1941, or ALA 1949. Use only for rare or archival materials.	
035 (See 5.2.5 for more details.)	The form of OCLC numbers has changed over many years. Records sent from local systems may contain various forms of OCLC control numbers, which may cause errors in matching, de-duplicating, and linking between OCLC systems and local systems. Or Records may contain no OCLC control numbers.	When you ordered your project, you entered the location and format of the OCLC control numbers in your records. If the format is inconsistent, you can work with the database specialist to normalize the OCLC numbers (that is, make them consistent).	Yes

Type of common error	Description of error	Prevention	Preprocessing routine available?
6xx	Subject headings are coded as local with 2nd indicator of 4 or with 2nd indicator of 7, and the source code is noted in ‡2.	<p>6xx 2nd indicator 4 Source not specified. The formulation of the subject added entry conforms to a controlled list, but the source cannot be specified by one of the thesaurus or subject heading systems covered by the other 2nd indicator values or by a code for a specific subject heading list in ‡2.</p> <p>6xx 2nd indicator 7 plus ‡2 Source is specified in ‡2. Subject headings or terms are based on other subject authorities (i.e., on authorities other than those listed here). Identify the source in ‡2.</p>	

5.2.5 More about the 035 field and OCLC control numbers

Batchload uses OCLC control numbers in your records in the following ways:

- To match WorldCat records.
- In local systems to link local records to WorldCat records.
- In batchload preprocessing to de-duplicate the records you submit for processing.

Because the format of OCLC control numbers has changed over many years and many local systems remove it or put it in a field other than the 035, OCLC batch services uses preprocessing routines to map existing OCLC control numbers to field 035 ‡a, prefixed with (OCoLC).

The following table describes the changes in OCLC control numbers formatting and location over time. Records can contain any of these formats and locations for the OCLC number.

Description	Format	Examples
In records created 12 November 2006 and before, field 001 contains the OCLC control number	OCLC numbers 1 through 99999999: Has an ocm prefix Composed of 8 digits, right justified, with leading zeros	ocm00012345
In records created after 12 November 2006, field 035 contains the OCLC control number	OCLC numbers 100,000,000 and higher: Has an ocn prefix. Composed of 9 digits.	ocn123456789

Description	Format	Examples
Sometime after November 2006, the OCLC control number in field 035 Ꞥa is prefixed by (OCoLC) .	OCLC number is a variable-length numeric string with no leading zeros.	(OCoLC)198765401

Notes:

- Field 019 contains the OCLC control number of a deleted or merged record replaced by the current one.
- If located in field 001, the OCLC control number is prefixed by **ocm** or **ocn**.

OCLC control numbers in records you send for batchload

If your records have the OCLC control number in field 035 Ꞥa with the **(OCoLC)** prefix, it will be used for matching.

If your records contain no OCLC numbers at all, Batch Services uses extended matching and/or other unique search keys to match the records with WorldCat records.

If your records contain different location/format of OCLC control numbers, Batch Services moves them to 035 Ꞥa for matching.

Note: If you ordered your own records returned with OCLC control numbers merged into the records (in the question, “What output do you want to receive when your project is completed?”), the question also lets you choose the field/subfield where you want to locate the merged number and its format.

After you select “Cross Reference Report and your records,” a new section labeled “OCLC control number and location” drops down, and you can select one of these locations for the OCLC number:

001, **035 Ꞥa**, or **Other** (you specify a field)

You can also specify one of these formats:

(OCoLC)1234, **(OCoLC)ocm00001234**, **ocm00001234**, **00001234**, or **1234**

See more [information about OCLC control numbers](#) on the OCLC Web site.