

Appendix B: Merging Specifications

Last revised September 2010

- B.1 Merging as part of batchload.....2
- B.2 Merging process3
- B.3 Mergeable fields.....4
- B.4 Table of mergeable fields and rules5
- B.5 Merge type rules7
- B.6 Standard Merge Rules8
- B.7 Field Specific Merge Notes9

B.1 Merging as part of batchload

Merging is one of a suite of software applications used to coordinate holdings and maintain records in the WorldCat database. This appendix explains merging of incoming files of records from local library systems for batchload.

Merging is used in conjunction with other applications and systems, all of which affect record output, master records, and archive records. The diagram below shows merging as part of the batchload process:

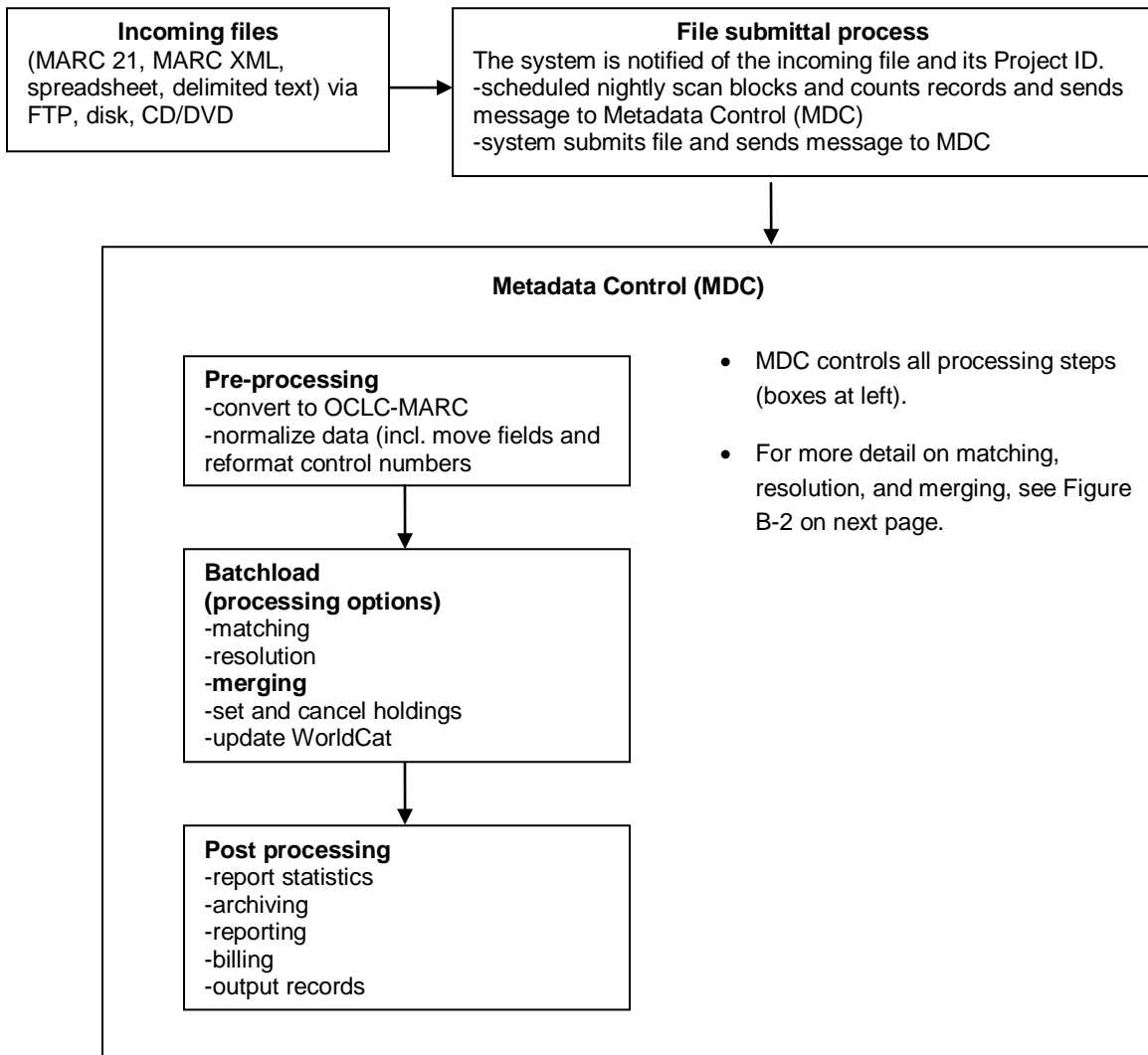


Figure B-1: Overview of merging in batchload

B.2 Merging process

Merging takes place after matching has identified one or more WorldCat records that match the incoming record. Resolution software determines which of the records (the incoming record or one of the matching WorldCat records) is the most complete and designates it the *Retain* record. In batchload, the library's record is usually designated the *Replace* record and the best WorldCat record is designated the Retain record. Other matching WorldCat records are ignored. Next, the selected mergeable fields in the Replace record are copied into the Retain record. Then, Merge software performs a field-by-field merge, with the fields chosen according to the rules outlined in this appendix. After merging, the Retain record is put back into WorldCat as the Master record.

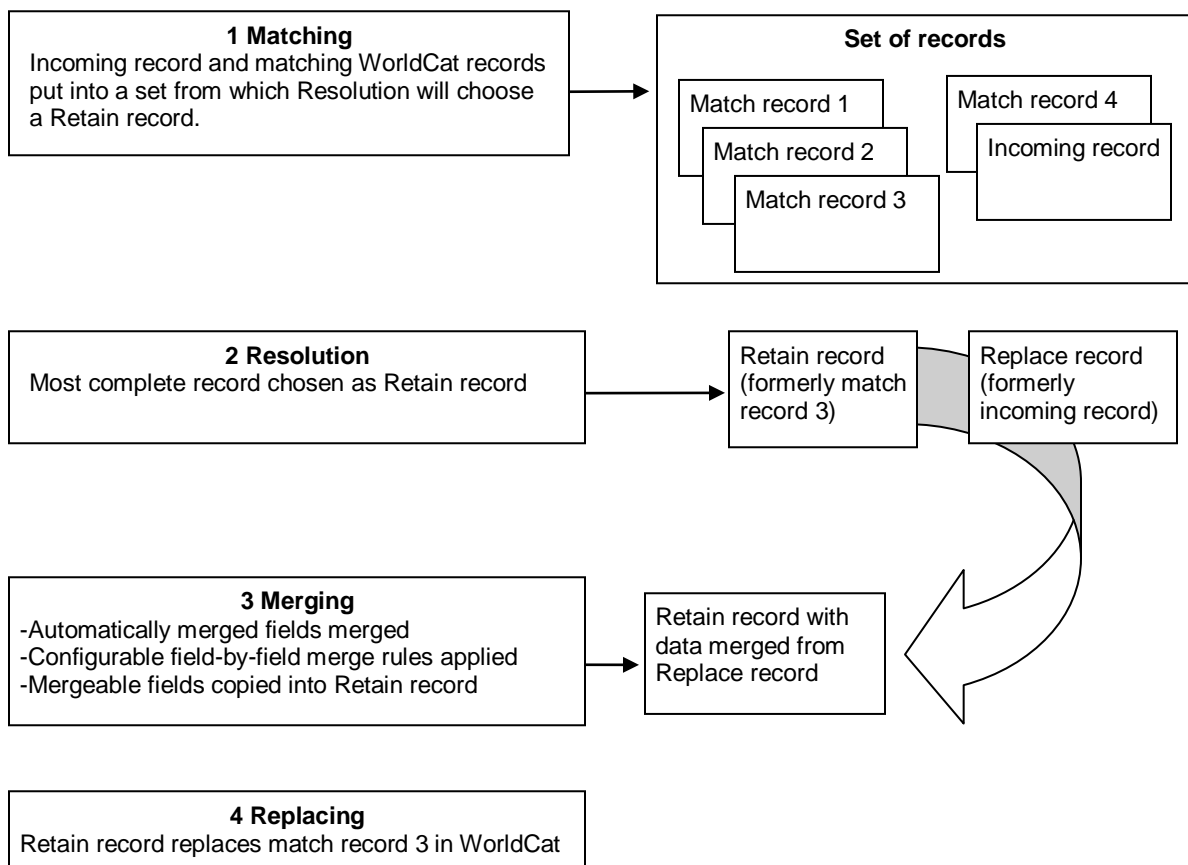


Figure B-2: Overview of matching, resolution, and merging

B.3 Mergeable fields

Merging is performed by a system called metadata capture (MDC), which controls data processing for individual projects and files. Merging can be turned on or off. If merging is off, no merging takes place. If merging is on, there are two kinds of mergeable fields:

- Automatically-merged fields:

029
040
938

In addition to the above fields, the information from controlled heading fields (1xx, 4xx, 6xx, 7xx) that indicates a given field is authority controlled is also automatically merged.

- Configurable fields are selected from the table below by the database specialist in consultation with the library:

006	033	070	306	583	776
007	037	072	504	586	800-830
010	040	074	505	600-651	856
015	041	080	506	653	880
016	043	082/092	508	654	887
020	045	083	511	655	891
022	047	084	520	656	936
024	048	085	526	657	989
027/088	050/090	086	530	658	
028	052	088	538	662	
030	055	258	542	751	
031	060/096	300/305	546	753	

Note: Batch Services policy is to merge some configurable fields if the records would be enhanced. For example, 050, 082, or 6xx fields with second indicator 0 are merged to retain any enhancement.

Note: Since the library's record is usually not chosen as the Retain record, fields are usually merged from it into the master record in WorldCat.

B.4 Table of mergeable fields and rules

This table lists each mergeable field (automatic and configurable) and the rules that apply to it. A description of each Merge type, rule, and notes follow the table.

Merge Table									
Field	Rules								Note
	Merge Type	CONSER	NLM	Format	Lang. of Cat	Bib lvl	Uniqueness	Non-Latin	
006	See Note.	X					X		Note 006
007	3	X		X					
010	2	X		X					
015	8			X			X		Note 015
016	8		X			X	X		Note 016
020	6	X		X		X	X	X	Note 020
022	2	X		X		X			
024	8			X		X	X	X	Note 024
027/088	See Note.			X		X			Note 027/088
028	7			X		X		X	
029	8						X		Note 029
030	6			X		X			
031	7	X		X				X	
033	3	X		X					
037	7	X		X		X		X	
040	See Note.						X		Note 040
041	3	X		X			X		
043	2	X		X				X	
045	2	X		X					
047	2	X		X					
048	3	X		X					
050/090	5	X		X					Note 050/090
052	3	X		X			X	X	Note 052
055	7	X		X				X	
060/096	0/1	X	X	X					
070	7	X		X					
072	7	X		X			X		Note 072
074	7	X		X					
080	8	X		X			X		Note 080
082/092	5	X		X			X		Note 082/092
083	7	X					X	X	Note 083
084	8	X		X			X	X	Note 084
085	7	X					X	X	Note 085
086	8	X		X			X	X	Note 086
088				X		X		X	Note 027/088

Merge Table									
Field	Rules								Note
	Merge Type	CONSER	NLM	Format	Lang. of Cat	Bib lvl	Uniqueness	Non-Latin	
258	7	X		X				X	
300/305	2	X		X	X	X		X	Note 300/305
306	2	X		X		X			
504	7	X		X	X	X		X	
505	7	X		X		X		X	
506	8	X		X		X	X	X	Note 506
508	7	X		X	X	X		X	
511	7	X		X	X	X		X	
520	7	X		X	X	X		X	
526	8	X		X	X	X	X		Note 526
530	7	X			X	X		X	
538	7	X		X	X	X		X	
542	8	X						X	
546	7	X		X	X	X		X	
583	8	X		X	X	X		X	
586	7	X		X	X	X	X	X	
600-651	4		Note NLM Rule	X			X	X	Note 600-651
653	7			X				X	
654	7			X			X	X	Note 654
655	7			X			X	X	Note 655
656	7			X			X	X	Note 656
657	7			X			X	X	Note 657
658	7			X			X	X	Note 658
662	7			X			X	X	Note 662
751	7	X					X	X	Note 751
753	7	X		X	X			X	
776	7	X		X		X	X	X	
800-830	See Note	X			X	X		X	Note 800-830
856	8	Note 856					X	X	Note 856
880	See Note	X							See Non-Latin Rules (page 13)
887	8	X		X			X		Note 887
891	7			X		X			
936	See Note							X	Note 936
938	8	X					X		Note 938
989	2	X		X					

B.5 Merge type rules

Merge Type 0 and Merge Type 1: Merge Type rules 0 and 1 apply to 060 and 096 fields and work in conjunction with the NLM Rule. For the NLM rule, see page 9.

Merge Type 0

- Merge Type 0 is applied for call numbers from NLM.
- Field 060 is retained over field 096. Field 096 is retained only if there are no 060 fields.
- If the incoming record is in an NLM project, its 060 is always the retained, even when one of the replace records is also from NLM.
- For non-NLM projects, if one of the records in a set of matching records is NLM (determined by presence of 029 with value *NLM*), the NLM 060 is always retained, regardless of which record is selected as the retain record.

Merge Type 1

- Merge Type 1 is applied for call numbers not from NLM.
- If none of the call numbers are from an NLM record, all occurrences of either the 060 or 096 from the first replacement record in the set are retained.
- If any record in the set includes field 060 with second indicator other than 4, the 096s are not retained

Merge Type 2

If the field is not present in the retain record, select the first occurrence of this field from the replace record.

Merge Type 3

The combination of the field and the indicator values determine if the fields merge.

- If the indicators are different, the field merges.
- If the field is not present in the retain record, the field merges.
- If the field is present in both records and the indicators are the same, the field does not merge.

Merge Type 4

See Note 600-651 below.

Merge Type 5

See Notes 050/090 and 082/092 below.

Merge Type 6

Merge Type 6 applies to field 020 and 030.

Field 020: See Note 020 below.

Field 030: Each field 030, subfield \$a in the replace record is compared with each 030 subfields \$a and \$z in the retain record. If subfields \$a and \$z are unique, they are added to the retain record.

Merge Type 7

If the field is not in the retain record, merge all occurrences of it from the first record in the replace record set that contains the field. If the field is already present in the retain record, apply the uniqueness test as described in the notes for each of the fields (See Table of mergeable fields and rules above).

Merge Type 8

If the field is not in the retain record, merge all occurrences of it from the replace record set. If the field is already present in the retain record, apply the uniqueness test as described in the notes for each of the fields (See Table of mergeable fields and rules above).

B.6 Standard Merge Rules

CONSER rule: The field will not merge into a CONSER record. A CONSER record has Bib lvl s and field 042 containing one of these codes:

<i>isds/c</i>	<i>lcac</i>	<i>lcd</i>	<i>nlc</i>	<i>nst</i>	<i>premarc</i>
<i>lc</i>	<i>lccopycat</i>	<i>msc</i>	<i>nsdp</i>	<i>pcc</i>	

NLM rule: OCLC has an agreement with NLM to always retain the NLM version of medical data. The NLM rule applies to medical information in fields 016, 060, 096, and MeSH subject headings (600-651 with second indicator 2). If an incoming NLM record matches a record that contains these fields, the NLM fields override the corresponding fields in the existing record. For example, if an LC record is the retain record and NLM the replace record, NLM's 016, 060/096 and MeSH subject headings will override the 016, 060/096 and MeSH subject headings in the LC record.

Type/BLvl Rule. The field will not merge if the Type/BLvl combinations are not compatible. For the valid combinations, see the Type section of *Bibliographic Formats and Standards*: <http://www.oclc.org/bibformats/en/field/type.shtml#combinationtypeblvl>.

Language of Cataloging Rule. The field will not merge if the Language of Cataloging (040 subfield \$b) in either the retain or replace record is different from the language of cataloging in the other record(s). If 040 subfield \$b is missing from any record, it is treated as if it were *eng* (English).

Bibliographic Level Rule (BLvl). The field will not merge if the bibliographic level in the retain record is different from the bibliographic level in the replace record.

Uniqueness Test. A uniqueness test determines if the data in a subfield is unique or the same. Unique data is merged into the retain record. If the data is the same, the field is not merged. If a field in the Merge Table has a uniqueness test, it is described in B.7, Field specific merge notes below.

Non-Latin Rules

In consultation with the Batchload Specialist, libraries can choose from four options for merging non-Latin data:

1) None. Field 880 will not merge.

2) Merge Latin only. The usual merge rules apply; if a field is selected to merge, it will merge. However, if there is non-Latin text in a linked 880 field, the 880 will not merge. The subfield \$6 in the Latin-only field will be deleted. This option is selected only if the non-Latin text and field are invalid.

3) Merge Latin and Linked Non-Latin. If a Latin-only field is going to be merged and it is linked to field 880, both fields will be merged. Field 066 in the retain record will be set accordingly. The occurrence number in subfield \$6 is recalculated. Unlinked non-Latin fields are not merged, even if they include data that would otherwise be a candidate for merging.

4) Merge Latin, Linked Non-Latin and Unlinked Non-Latin. Includes all of the features of *3 Merge Latin and Linked Non-Latin* options with the following additions:

- If the retain record contains a Latin-only field for which there is a non-Latin field in the replace record, the Latin-only text is compared to the Latin-only text in the same field in the replace record. If it matches, the non-Latin text is merged and the fields are linked.
- If it does not match, the non-Latin text is merged to the retain record as an unlinked 880 field. Field 066 is created or modified if appropriate.

Special rules for Series (fields 490, 800-830). For option 4 above, the Latin-only text in the replace record is compared to all of the series fields in an attempt to link to one of them. If any series field matches, the non-Latin text is merged to the retain record. The Non-Latin field is added to retain record as a new field with the same field number as the field it matched.

Special rules for Name heading fields (100-130, 700-730). For option 4 above, the Latin-only text in the replace record is compared to equivalent name heading fields in an attempt to link to one of them. If any name heading field matches, the non-Latin text is merged to the retain record. The Non-Latin field is added to retain record as a new field with the same field number as the field it matched.

Equivalent name heading fields:

100 is equivalent to 700

110 is equivalent to 710

111 is equivalent to 711

130 is equivalent to 730

B.7 Field Specific Merge Notes

Note 006—The 006 field merges under any of these conditions:

- The code in the first position of field 006 does not match the code in any other 006
- The code in the first position of field 006 does not match the Leader/6 value
- If the 006/00 is s and Leader/07 is s, field 006 will not merge

Note 015—The uniqueness test is applied to subfield \$2; if it is unique, field 015 merges. If they are the same, the contents of subfield \$a are compared; if the subfields contain unique data, they are retained.

Note 016—If one of the 016's is in an NLM record (040 subfield \$c *NLM*), it is always retained. The uniqueness test is applied to the first indicator and subfield \$2; if they are unique, they are retained. If they are the same, subfields \$a and \$z are compared; if subfields \$a and \$z contain unique data, field 016 is retained.

Note 020—Uniqueness test is applied to subfields \$a and \$z. If they are different, they are retained, along with subfield \$b.

Note 024—The uniqueness test is applied to the first indicator and subfield \$2; if they are unique, they can be merged. If they are the same, the contents of subfields \$a and \$z are compared; if the subfields contain unique data, field 024 is retained.

Note 027/088—If either field 027 or field 088 (or both) already exists in the retain record, do not merge any 027 or 088 fields.

Note 029—The uniqueness test is applied to the concatenation of subfields \$a, \$b, \$c and \$t. If the concatenated text is different, field 029 will merge.

Note 040—Subfields \$c and \$d from the replace record are merged into the retain record in 040 subfield \$d unless the symbol is already present in field 040. **Exceptions:** These symbols are not merged: OCL, OCLCQ, OCLCA, OCLCG.

Note 050/090—There is a hierarchy of decisions considered prior to selecting the data to retain.

- If present, a call number with class letters is retained and call number without class letters is not.
- The LC-assigned class number is always retained when present in the set of records; it is identified by the 040 subfield \$c *DLC* or by the second indicator value 0. If there is more than one LC-assigned class number in a record, all are retained.
- If there is no LC-assigned class number, retain any single 050 field if present. If no 050 field is present, retain field 090.

When field 090 is retained, it is converted to field 050 with a second indicator value 4.

Note 052—The uniqueness test is applied to subfield \$2. If the subfields \$2 are unique, field 052 will merge.

Note 072—The uniqueness test is applied to the second indicator. If the second indicators are unique, field 072 will merge. The only instance of second indicators being the same that might result in a merge is if the second indicators are 7, in which case subfield \$2 is uniqueness tested. If the subfields \$2 are unique, then field 072 will merge.

Note 080—The uniqueness test is applied to subfield \$2. If the subfields \$2 are unique, they are merged. If the subfields \$2 are the same, the subfields \$a are uniqueness tested. If the subfields \$a are unique, they are retained.

Note 082/092—There is a hierarchy of decisions considered prior to selecting the data to retain.

- If the retain record has an 082 or 092 beginning with alphabetic characters or lacking numeric characters, it is deleted; exceptions are *Fic*, *E*, and *B* which are considered valid DDC numbers and are retained.
- The numbers in subfield \$a are compared. If the numbers are the same, preference is given to the 082 field with second indicator 0 (meaning, "assigned by LC").
- If the numbers in subfield \$a are the same and none is from LC, retain the one with subfield \$2.

- If the numbers in subfield \$a and subfield \$2 are the same, retain the one in the retained record.
- If the numbers in subfield \$a and subfield \$2 are the same and are not in the retained record, retain the longest 082 or 092 field.

When field 092 is retained, it is converted to field 082 with a second indicator value 4 and subfields \$e and \$f are deleted.

Note 083—The uniqueness test is applied to the contents of subfield \$2. If the subfields \$2 are unique, field 083 will merge.

Note 084—The uniqueness test is applied to the contents of subfield \$2. If the subfields \$2 are the same, the uniqueness test is applied to subfields \$a and \$z. If subfields \$a and \$z are unique, they are retained.

Note 085—The uniqueness test is applied to the contents of subfield \$2. If the subfields \$2 are unique, field 085 will merge.

Note 086—The uniqueness test is applied to the first indicator and the contents of subfield \$2. If the first indicator and subfield \$2 code are unique, the field is merged. If the first indicator and subfield \$2 code are the same, the uniqueness test is applied to subfields \$a and \$z. If subfields \$a and \$z are unique, field 086 is retained.

Note 300/305—If either record has multiple fields 300 or 305, no merge occurs. Fields 300/305 are candidates for replacing only under the following conditions:

- Subfield \$a contains 'p. cm.'
- Encoding Level is 8 and the field contains only subfield \$a
- Record has 040 subfield \$c *UKM* and 300/305 subfield \$a begins with [

If the retain record has no 300 or it meets one of the criteria above, it can be replaced by a 300 from the first record merged to it.

Note 506—The uniqueness test is applied to subfield \$5. If subfields \$5 are unique or absent, field 506 will merge.

Note 526—The uniqueness test is applied to a combination of subfields \$a and \$5. If the combination is unique or subfield \$5 is absent, field 526 will merge.

Note 600-651—The uniqueness test is applied to the second indicator. If the second indicators are unique, fields 600-651 will merge. The only instance of second indicators being the same that might result in a merge is if the second indicators are 7, in which case subfield \$2 is uniqueness tested. If the subfields \$2 are unique, then fields 600-651 will merge. All 600-651 fields with the same indicator value (or subfield \$2 as appropriate) are merged or not as a group (for example, if one is retained all are retained). **NLM Exception:** If one of the records is an NLM record (040 subfield \$c *NLM*), NLM subject headings (indicator value 2) always override other subject headings with indicator value 2.

Note 654—The uniqueness test is applied to subfield \$2. If the subfields \$2 are unique, field 654 will merge.

Note 655—The uniqueness test is applied to the second indicator. If the second indicators are unique, field 655 will merge. The only instance of second indicators being the same that might result in a merge is if the second indicators are 7, in which case subfield \$2 is uniqueness tested. If the subfields \$2 are unique, field 655 will merge.

Note 656—The uniqueness test is applied to the second indicator. If the second indicators are unique, field 656 will merge. The only instance of second indicators being the same that might

result in a merge is if the second indicators are 7, in which case subfield \$2 is uniqueness tested. If the subfields \$2 are unique, field 656 will merge.

Note 657—The uniqueness test is applied to the second indicator. If the second indicators are unique, field 657 will merge. The only instance of second indicators being the same that might result in a merge is if the second indicators are 7, in which case subfield \$2 is uniqueness tested. If the subfield \$2's are unique, field 657 will merge.

Note 658—The uniqueness test is applied to subfield \$2. If the subfields \$2 are unique, field 658 will merge.

Note 662—The uniqueness test is applied to subfield \$2. If the subfields \$2 are unique, field 662 will merge.

Note 751—The uniqueness test is applied to the contents of subfield \$2. If the subfields \$2 are unique, field 751 will merge.

Note 800-830—All series fields (490, 800, 810, 811, and 830) are treated as a group. Series information is retained based on these rules:

- Traced series (800-830) are always retained over untraced series (490 with first indicator 1).
- If there are multiple replace records with traced series, choose the first one merged.
- Non-Latin rules: See *Non-Latin Rules* section on page 9.

Note 856—The uniqueness is applied to subfield \$u. If the subfields \$u are unique, field 856 merges. For CONSER records, 856 does not merge unless it meets one of these criteria:

- If field 856 subfield \$u contains the text "books.google.com", field 856 merges if it is not already present.
- If field 856 subfield \$u contains the text "firstsearch.oclc.org", field 856 merges if it is not already present.
- If field 029 subfield \$a contains the text "CNTNT", field 856 merges if it is not already present.

Note 887—The uniqueness test is applied to subfield \$2. If the subfields \$2 are unique or absent, field 887 is transferred.

Note 936—If field 936 does not start with *PR*, it is not retained. If field 936 begins with *PR*, it is a merge type 2. OCLC numbers from all of the 936 fields that begin with *PR* are concatenated into a single string. Duplicate OCLC control numbers are deleted. The OCLC control numbers are compared to the 001 and 019 subfield \$a of the retain record. If the number in field 936 is in fields 001 or 019, it is deleted from the replace record. If all OCLC control numbers are deleted, field 936 is also deleted from the replace record.

Note 938—The uniqueness test is applied to subfield \$b. If the subfields \$b are identical in both records, the 938 will be retained from the incoming record.