

DRAFT Recommendation from the SDTF  
9/28/2005

*Recommendation  
approved 11/15  
Murphy/Marell  
forwarded to  
Sirsi/Dynix*

## **Proposed Initial Matching Overlay Loading Algorithm for Merge Database**

### **Rationale:**

In determining specifics for the overlay loading algorithm, the Single Database Task Force based its recommendations on the following:

- Retaining accessibility of unique local information on bibliographic records.
- Ensuring records are not overlaid unless a minimum of at least 2 specified match points exist between records to be merge
- Current MARC Standards, available online from: <http://www.loc.gov/marc/>
- OCLC Tape Loading Guide, Section 7: How Records are Matched.
- INTEROFFICE MEMORANDUM: Sunrise Bibliographic Matching Algorithm & Record Consolidation, from DALNET Union Catalog/Authorities Task Force, 1999
- Documentation: Updated Algorithm logic for master record, matching, and action of MARC load program, from Sam Foster, of SirsiDynix, 1998

The Single Database Task Force examined all repeatable MARC fields\* and used the following criteria in determining which fields should be repeated and separate entries retained on the merged records in the database:

- Examination of member databases syntax and validation to verify if fields were currently being used. It was felt that if no library was currently using a particular MARC field then that field did not need special retention in the initial merge.
- Necessity of retaining specific local information for staff access
- Necessity of retaining specific local information for public access.
- Conference with DALNET Project Managers and anyone whom they wished to consult, i.e. their catalogers, on impact for local/member practices and/or specific fields to retain for each institution.

\*A complete list of repeatable MARC fields is available online from:  
<http://www.loc.gov/marc/bibliographic/ecbdhome.html>

Only unique entries in repeatable fields will be retained; multiple entries for repeatable fields that contain the same information will be overlaid. All other non-repeatable MARC fields will be overlaid when merging files, thus the significance of the order in which records are added to the merge database from each institution

### **Note:**

All recommendations for the merge algorithm are based on the assumption that the algorithm will be changed, tweaked, altered, etc. if desired results are not achieved once in the testing environment, with the intent that the final algorithm be perfected before a final production database is created. The recommended algorithm is a best plan beginning based on the above described rationale.

### **Proposed Match Point Logic for Record Overlay:**

Match point logic is applied after record is read from file and before any record is loaded. Match points in order of precedence. Later referred to by number.

1. Exact match on OCLC (001,019, 035)
2. Exact Match on ISBN (020a)
3. Exact match on ISSN (022a)
4. Exact match on Government Documents Number (086a)
5. Exact Match on LCCN (010a)
6. Match on first five words (245a) and (245h)

```
IF (1 AND (2 OR 3 OR 4 OR 5 OR 6)) THEN
  RETURN TRUE (match)
ELSE IF (any two matches on 2-6) THEN
  RETURN TRUE (match)
ELSE
  RETURN FALSE (no match)
END
```

### **Proposed Repeatable Fields to retain Multiple Entries for in Record Consolidation:**

Only unique entries in repeatable fields will be retained; multiple entries for repeatable fields that contain the same information will be overlaid. All other non-repeatable MARC fields will be overlaid when merging files, thus the significance of the order in which records are added to the merge database from each institution.

Fields 010-04X contain control and linking numbers, standard numbers, and codes that relate to the bibliographic item described in the record.

- 013 - PATENT CONTROL INFORMATION (R)
- 015 - NATIONAL BIBLIOGRAPHY NUMBER (R)
- 016 - NATIONAL BIBLIOGRAPHIC AGENCY CONTROL NUMBER (R)
- 017 - COPYRIGHT OR LEGAL DEPOSIT NUMBER (R)
- 020 - INTERNATIONAL STANDARD BOOK NUMBER (R)
- 022 - INTERNATIONAL STANDARD SERIAL NUMBER (R)
- 024 - OTHER STANDARD IDENTIFIER (R)
- 025 - OVERSEAS ACQUISITION NUMBER (R)
- 026 - FINGERPRINT IDENTIFIER (R)
- 027 - STANDARD TECHNICAL REPORT NUMBER (R)
- 028 - PUBLISHER NUMBER (R)
- 030 - CODEN DESIGNATION (R)
- 031 - MUSICAL INCIPITS INFORMATION (R)
- 032 - POSTAL REGISTRATION NUMBER (R)
- 033 - DATE/TIME AND PLACE OF AN EVENT (R)
- 034 - CODED CARTOGRAPHIC MATHEMATICAL DATA (R)

- 035 - SYSTEM CONTROL NUMBER (R)
- 037 - SOURCE OF ACQUISITION (R)
- 041 - LANGUAGE CODE (R)
- 046 - SPECIAL CODED DATES (R)
- 048 - NUMBER OF MUSICAL INSTRUMENTS OR VOICES CODE (R)

Fields 210-24X contain the title of the item described in the bibliographic record and variant titles that also apply to the item. These fields may be used to generate access points and to display notes for the various titles.

- 210 - ABBREVIATED TITLE (R)
- 222 - KEY TITLE (R)
- 242 - TRANSLATION OF TITLE BY CATALOGING AGENCY (R)
- 246 - VARYING FORM OF TITLE (R) (add |5 notes)
- 247 - FORMER TITLE (R)

Fields 250-270 contain edition, imprint, address, and other publication source information, and data related to specific forms of material that apply to the item described in the bibliographic record.

- 255 - CARTOGRAPHIC MATHEMATICAL DATA (R)
- 258 - PHILATELIC ISSUE DATA (R)
- 260 - PUBLICATION, DISTRIBUTION, ETC. (IMPRINT) (R)
- 270 - ADDRESS (R)

Fields 300-362 contain physical characteristics, publication frequency, price, and physical arrangement information for the item described in the bibliographic record. Information about the dissemination of bibliographic items and the security status of bibliographic data relating to them is also recorded in fields in this group.

- 300 - PHYSICAL DESCRIPTION (R)
- 307 - HOURS, ETC. (R)
- 321 - FORMER PUBLICATION FREQUENCY (R)
- 340 - PHYSICAL MEDIUM (R)
- 342 - GEOSPATIAL REFERENCE DATA (R)
- 343 - PLANAR COORDINATE DATA (R)
- 351 - ORGANIZATION AND ARRANGEMENT OF MATERIALS (R)
- 352 - DIGITAL GRAPHIC REPRESENTATION (R)
- 355 - SECURITY CLASSIFICATION CONTROL (R)
- 362 - DATES OF PUBLICATION AND/OR SEQUENTIAL DESIGNATION (R)
- 365 - TRADE PRICE (R)
- 366 - TRADE AVAILABILITY INFORMATION (R)

Fields 440 and 490 contain series statements. Field 440 also provides an added entry for the series. A series statement is recorded in field 490 when either it is given no added entry in the record (an untraced series) or is given an added entry in a form that differs

from the series statement appearing on the item. One of fields 800-830 (Series added entries) is used in conjunction with field 490 when the series is to be traced differently.

The parentheses that customarily enclose a series statement and the tracing *Series.*, preceded by a roman numeral, in some displays for field 440 are not carried in the MARC record. They may be generated based on the field tag.

- 440 - SERIES STATEMENT/ADDED ENTRY--TITLE (R)
- 490 - SERIES STATEMENT (R)

Fields 500-53X contain notes relating to aspects of bibliographic items that are not specific to any particular type of material or control (e.g., archival control). Specialized note fields are used when access to the data is needed and/or the note is introduced by a distinctive word or phrase. Field 500 is used for all other note information.

- 500 - GENERAL NOTE (R)
- 501 - WITH NOTE (R)
- 505 - FORMATTED CONTENTS NOTE (R) (add |g notes)
- 506 - RESTRICTIONS ON ACCESS NOTE (R) (add|5 notes)

**After checking with Dynix it is noted that it will not be possible to retain just the longest 505 field as had initially been proposed by the Project Managers.**

- 510 - CITATION REFERENCES NOTE (R)
- 520 - SUMMARY, ETC.(R)
- 521 - TARGET AUDIENCE NOTE (R) (add |b notes)
- 530 - ADDITIONAL PHYSICAL FORM AVAILABLE NOTE (R)
- 533 - REPRODUCTION NOTE (R)
- 534 - ORIGINAL VERSION NOTE (R)
- 535 - LOCATION OF ORIGINALS/DUPLICATES NOTE (R)

**It is not planned to retain multiple entries for any of these fields but Project Managers will be consulted to see if any of these fields have been used by any libraries. If any of these fields have been used by more than one library they will be retained as multiple entries.**

Fields 53X-58X contain bibliographic notes which are more specific in nature. A significant number of these note fields are used to record information relating to the archival or continuing control of the bibliographic item.

- 538 - SYSTEM DETAILS NOTE
- 541 - IMMEDIATE SOURCE OF ACQUISITION NOTE (R)
- 563 - BINDING INFORMATION (R) (add |5 notes)
- 586 - AWARDS NOTE (R)
- 59X - LOCAL NOTES (R) (add |z notes)

**If these fields (above) are used, we plan on retaining multiple entries for them.**

- 565 - CASE FILE CHARACTERISTICS NOTE (R)

**It is not planned to retain multiple entries for this field but Project Managers will be consulted to see if this field has been used by any libraries. If this field has been used by more than one library it will be retained as multiple entries.**

**In general, it is planned to check the syntax and validation tables in Horizon to see how many of the 5xx fields have been utilized and for those that might have been, it is planned that reports will be generated on which fields have been used in which databases so that members may give direction for the multiple retention or overlay of specified fields.**

Fields 600-65X (with the exception of field 653 that is used for uncontrolled index terms) contain subject headings or access terms that provide additional access to a bibliographic record through a heading or term that is constructed according to established subject cataloging or thesaurus-building principles and guidelines. The standard list or authority file used is identified by the value in the second indicator position or by the MARC source code contained in subfield \$2 that is used in conjunction with value 7.

The Arabic number that precedes a subject access field in some displays is not carried in the MARC record. It may be generated based on the field tag. The dash ( -- ) that precedes a general subject subdivision is not carried in the MARC record; it may be generated based on the presence of subfields, \$v, \$x, \$y, and \$z.

A name portion containing open-ended date ends with a space when it is followed by other data in the same subfield; it does not end with a space when it is followed by another subfield.

- 600 - SUBJECT ADDED ENTRY--PERSONAL NAME (R)
- 610 - SUBJECT ADDED ENTRY--CORPORATE NAME (R)
- 611 - SUBJECT ADDED ENTRY--MEETING NAME (R)
- 630 - SUBJECT ADDED ENTRY--UNIFORM TITLE (R)
- 648 - SUBJECT ADDED ENTRY--CHRONOLOGICAL TERM (R)
- 650 - SUBJECT ADDED ENTRY--TOPICAL TERM (R)
- 651 - SUBJECT ADDED ENTRY--GEOGRAPHIC NAME (R)
- 653 - INDEX TERM--UNCONTROLLED (R)
- 654 - SUBJECT ADDED ENTRY--FACETED TOPICAL TERMS (R)
- 655 - INDEX TERM--GENRE/Form (R)
- 656 - INDEX TERM--OCCUPATION (R)
- 657 - INDEX TERM--FUNCTION (R)
- 658 - INDEX TERM--CURRICULUM OBJECTIVE (R)

Fields 700-75X contain a name and/or title or a term that provides access to a bibliographic record that is not provided through main entry (1XX), subject access

(6XX), series statement (4XX), series added entry (8XX), or title (20X-24X) fields. The roman numeral and the word *Title*: that precede an added entry field in some displays are not carried in the MARC record. They may be generated based on the field tag.

A name portion containing open-ended date ends with a space when it is followed by other data in the same subfield; it does not end with a space when it is followed by another subfield.

- 700 - ADDED ENTRY--PERSONAL NAME (R)
- 710 - ADDED ENTRY--CORPORATE NAME (R)
- 711 - ADDED ENTRY--MEETING NAME (R)
- 720 - ADDED ENTRY--UNCONTROLLED NAME (R)
- 730 - ADDED ENTRY--UNIFORM TITLE (R)
- 740 - ADDED ENTRY--UNCONTROLLED RELATED/ANALYTICAL TITLE (R)
- 752 - ADDED ENTRY--HIERARCHICAL PLACE NAME (R)
- 753 - SYSTEM DETAILS ACCESS TO COMPUTER FILES (R)
- 754 - ADDED ENTRY--TAXONOMIC IDENTIFICATION (R)

Fields 760-78X contain information that identifies other related bibliographic items. Each of the linking entry fields specifies a different relationship between the target item described in the record and a related item. These relationships fall into three categories: 1) related items that assist the user in continuing to search but are not physically required to obtain the target item (e.g., former entries for continuing resources, translations of the target item); 2) related items that have to be obtained physically in order to use the target item (e.g., the host item for a component part: a journal issue containing a specific article); 3) related items that are constituent units of a larger whole (e.g., the individual photographs contained in a visual material collection). The linking entry fields are intended to generate a note in a display of the record in which they appear; provide machine linkage between the bibliographic record for the target item and the bibliographic record for the related item, if the related item is covered by a separate record; and/or facilitate indexing.

- RECORD LINKS AND NOTES--GENERAL INFORMATION
- 760 - MAIN SERIES ENTRY (R)
- 762 - SUBSERIES ENTRY (R)
- 765 - ORIGINAL LANGUAGE ENTRY (R)
- 767 - TRANSLATION ENTRY (R)
- 770 - SUPPLEMENT/SPECIAL ISSUE ENTRY (R)
- 772 - SUPPLEMENT PARENT ENTRY (R)
- 773 - HOST ITEM ENTRY (R)
- 774 - CONSTITUENT UNIT ENTRY (R)
- 775 - OTHER EDITION ENTRY (R)
- 776 - ADDITIONAL PHYSICAL FORM ENTRY (R)
- 777 - ISSUED WITH ENTRY (R)
- 780 - PRECEDING ENTRY (R)
- 785 - SUCCEEDING ENTRY (R)

- 786 - DATA SOURCE ENTRY (R)
- 787 - NONSPECIFIC RELATIONSHIP ENTRY (R)

Fields 800-830 contain a name/title or a title used as a series added entry when the series statement is contained in field 490 (Series Statement) or field 500 (General Note) and a series added entry is required for the bibliographic record. No 8XX series added entry field is used in conjunction with a series statement in the 440 field because it functions both as part of the bibliographic description and as a series added entry.

The Roman numeral and the word *Series:* that precede a series added entry field in some displays are not carried in the MARC record. They may be generated based on the field tag.

A name portion containing open-ended date ends with a space when it is followed by other data in the same subfield; it does not end with a space when it is followed by another subfield.

- 800 - SERIES ADDED ENTRY--PERSONAL NAME (R)
- 810 - SERIES ADDED ENTRY--CORPORATE NAME (R)
- 811 - SERIES ADDED ENTRY--MEETING NAME (R)
- 830 - SERIES ADDED ENTRY--UNIFORM TITLE (R)

Fields 841-88X contain descriptions for data elements which are an integral part of the *MARC 21 Format for Bibliographic Data* and data elements which may appear either in bibliographic records or in separate MARC holdings records.

Field 850 is a bibliographic record field that contains minimal holdings location data. Field 852 contains more extensive location data. Field 856 covers information about the location of electronic resources. These fields are described in this document. Fields 841-845, 853-855, and 863-878 are only listed here. Fuller descriptions for them appear in the *MARC 21 Concise Format for Holdings Data*.

Fields 880, 886, and 887 are special-use bibliographic data fields.

- 843 - REPRODUCTION NOTE (R)
- 845 - TERMS GOVERNING USE AND REPRODUCTION NOTE (R)
- 850 - HOLDING INSTITUTION (R)
- 852 - LOCATION (R)
- 853 - CAPTIONS AND PATTERN--BASIC BIBLIOGRAPHIC UNIT (R)
- 854 - CAPTIONS AND PATTERN--SUPPLEMENTARY MATERIAL (R)
- 855 - CAPTIONS AND PATTERN--INDEXES (R)
- 856 - ELECTRONIC LOCATION AND ACCESS (R) (add |x notes)
- 863 - ENUMERATION AND CHRONOLOGY--BASIC BIBLIOGRAPHIC UNIT (R)
- 864 - ENUMERATION AND CHRONOLOGY--SUPPLEMENTAL MATERIAL (R)

- 865 - ENUMERATION AND CHRONOLOGY--INDEXES (R)
- 866 - TEXTUAL HOLDINGS--BASIC BIBLIOGRAPHIC UNIT (R)
- 867 - TEXTUAL HOLDINGS--SUPPLEMENTARY MATERIAL (R)
- 868 - TEXTUAL HOLDINGS--INDEXES (R)
- 876 - ITEM INFORMATION--BASIC BIBLIOGRAPHIC UNIT (R)
- 877 - ITEM INFORMATION--SUPPLEMENTARY MATERIAL (R)
- 878 - ITEM INFORMATION--INDEXES (R)
- 880 - ALTERNATE GRAPHIC REPRESENTATION (R)
- 886 - FOREIGN MARC INFORMATION FIELD (R)
- 887 - NON-MARC INFORMATION FIELD (R)

Much of the MARC information in this document is taken from: Current MARC Standards, available online from: <http://www.loc.gov/marc/>

[This document was approved by the DALNET Project Managers via online vote]