**Shared Bib Records with Multiple Series Fields**

**-- Issues and Suggested Solution**

Prepared by the Bibliographic Control and Discovery Subcommittee (BCDS), Multi-Series Cleanup Task Force: Annie Glerum (FSU), Ethan Fenichel (FAU), Elaine Dong (FIU). Last update, July 7, 2015.

Upon the request of BCDS, FLVC ran a report and pulled out 209,671 shared bib records with multiple series statement fields (440s and 490s).

***Problems with These Records***

* Obsolete 440 field(s) on records (please see the example 1 below)
* Multiple 490s/830s with same tracing: slightly different statements or numbering information in the $a subfield were all kept during the SB merge, and each 490/830 has different $5 institution code. (please see the example 2 and 3 below)
* Multiple 490 fields with different tracing: due to different tracing practice of SULs (please see the example 4 below)

***Issues for Indexing, Display, and Identification in Mango***

* If every 490 has a $5, then the field does not display in Union mango
* The series statement does not index correctly if the numbering is not correctly in the $v subfield
* The record should represent the same manifestation therefore the series statements should not differ

***Suggested Solution***

After some discussions at BCD meetings, we feel that instead of updating these records manually, a more efficient approach would be batch overlaying a select subset of records with OCLC master records to remove redundant/incorrect series statement fields.

***Sample Record Analysis***

We did the following analysis in order to identify what kind of records are fit to be overlaid and what’s the best approach to identify them:

Step 1

1,000 random sample records were pulled out (used MOD(ROW(),209)=1 in a macro so every 209th row was selected then copied/pasted as values) from Aleph, then we did 035 fields analysis and separated them into the following four categories:

1. Records with 035 OCLC# only (674 records, 67%): see the 1st sheet of the attached file sample035.xlsx
2. Records with both OCLC# and vendor# in 035 field (63 records, 6%): the majority of these records are identified as vendor records. If so, OCLC# needs to be flipped to 599. They can’t be updated by the method suggested in the statement.
3. CIS Microfiche Records (all have 035 gp#, 36 records, 3%): these records are for microfiche format but their OCLC records are for print format. They should not be overlaid by OCLC records.
4. Vendor records (no OCLC# in 035 fields, 285 records, 28%): these records can’t be updated by the method suggested in the statement. (see 2nd sheet of the attached file sample035.xlsx)

After discussion, we agreed that records in Group 2-4 can’t or shouldn’t be overlaid by their OCLC master records in order to remove redundant series fields.

Step 2

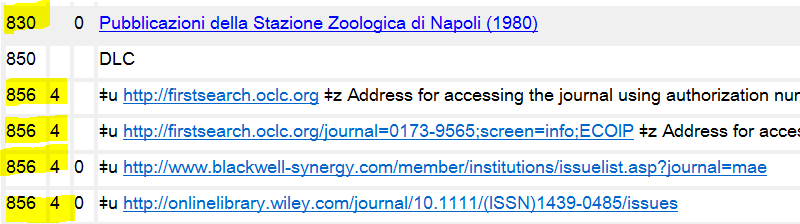
To examine more closely at the Group 1 records, we pulled out the 1st 10,000 records from Aleph (among 209,671 records with multiple series statement fields). Ethan created a script[[1]](#footnote-1), which could identify the formats of these 10,000 Aleph records:

|  |  |  |
| --- | --- | --- |
| No. of records with 035 OCLC# only | 7,535 | Percentage |
| Format: print | 6,697 | 89% |
| Format: electronic | 391 | 5% |
| Format: microform | 422 | 6% |
| Format: unknown | 25 | 0.3% |

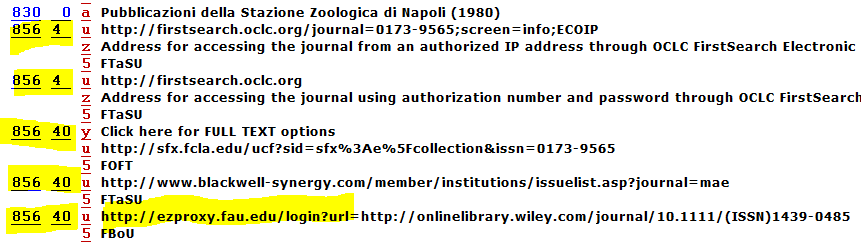
It shows the majority of Group 1 records (89%) are print format. There are also a small amount of electronic (5%), microform, and unknown format. We checked some records for each format and here are our findings:

1. Print format: they are safe to be overlaid
2. Electronic format: they should be excluded for the following reason:
3. Sometimes vendor records received from vendor and then loaded into Aleph are different from their OCLC master records;
4. An OCLC master record for electronic format usually contains several 856 fields/URLs from different vendors. If the OCLC record is exported to overlay the Aleph record, all these URLs on OCLC record will be added, and the 856 fields with $5 from Aleph records will be protected. See an example below:

OCLC record (#45205071)



Aleph record (sys# 020001910)



1. Microform format: OCLC records for a few of these are for print format.

In light of this finding, we think it is important to have the format comparison between OCLC and Aleph records. After the function was added to the program, a total of 17 records found a mismatched form between their OCLC and Aleph records.

1. Unknown format:

In order to identify the format of these records, Ethan added 007 field examination in addition to 008 field examination. It turns out most of them are map or GIS records. They are safe to be overlaid.

***Record Cleanup Plan***

Based on the findings from sample record analysis, we came up with the following steps to clean up these problematic records:

Step 1. Identify the subset of non-electronic records with 035 OCLC# only

1. Extract the MARC records from Aleph, 10,000 records a file
2. Run program to identify the Aleph records with 035 OCLC# only, and identify and exclude the electronic records
3. Extract the OCLC master records of the filtered records and save as mrc file
4. Run the program to compare the format between OCLC and Aleph records, and exclude the mismatched records

Step 2. Identify local series

Run program to compare the following fields:

1. Compare each 440 field to all the 490, and 830 fields of Aleph records. If found no match, then compare each 440 field to all the 490 and 830 fields of OCLC master records. They are identified as local series if no match found.
2. Compare each 490 field to all the 490 and 830 fields of Aleph records first, then to those of OCLC master records. They are identified as local series if no match found.
3. Compare each 830 field of Aleph records to the other 830 fields of Aleph records first, then to all the 830 fields of OCLC master records. They are identified as local series if no match found.

Step 3. Edit and protect local series fields

1. For each identified local series, change 440(s), 490(s), or 830(s) to 490/899 combination with $5 KEEPxx[[2]](#footnote-2) + $5 institution code.
2. Insert these local series fields to the OCLC mrc file using PyMARC.

Step 4. Overlay the subset of records by Genload

1. Create a profile for FLVC’s approval. The profile would set OCLC# as match point and protect the following fields:
2. Protect the following fields (with or without $5): 590, 599, 970, STA
3. Protect the following fields **with $5**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| MARC TAG | MARC TAG | MARC TAG | MARC TAG | MARC TAG |
| 351 | 541 | 690 | 830 | 909 |
| 500 | 545 | 691 | 845 | 951 |
| 501 | 561 | 699 | 856 |  |
| 506 | 562 | 790 | 896 |  |
| 520 | 563 | 791 | 897 |  |
| 533 | 583 | 796 | 898 |  |
| 540 | 584 | 797 | 899 |  |

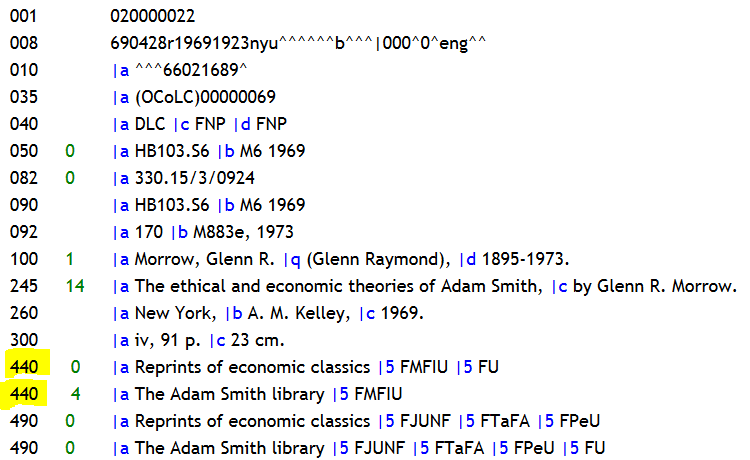
1. Protect the 655 \_7 field with the following $2

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| MARC TAG | 655 \_ 7 | 655 \_ 7 | 655 \_ 7 | 655 \_ 7 | 655 \_ 7 | 655 \_ 7 | 655 \_ 7 |
| **$2** | rbprov | rbbin | rbgenr | rbpap | rbpri | rbpub | rbtyp |

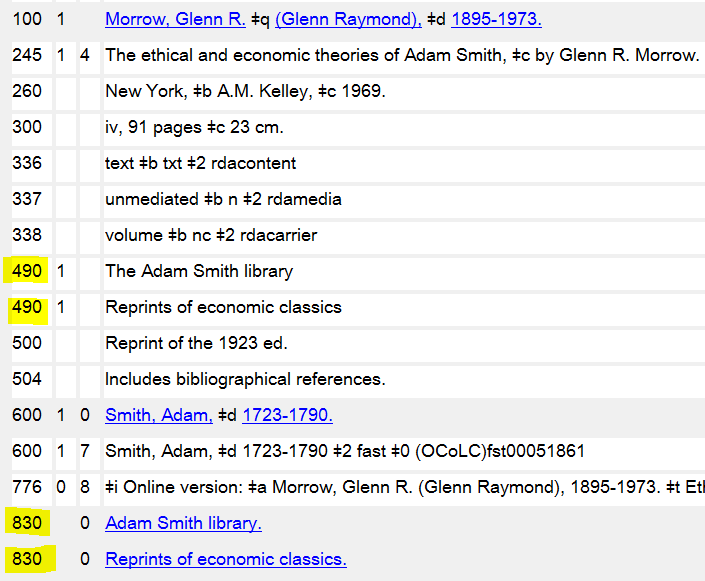
1. Use Genload to overlay the subset of Aleph records identified by Step 1

The multiple matches found during the loading process will not be overlaid. Duplicate CIS microfiche records would be caught by multi-matches.

**Example 1: Obsolete 440 fields on Aleph Record**



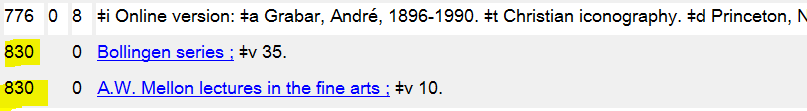
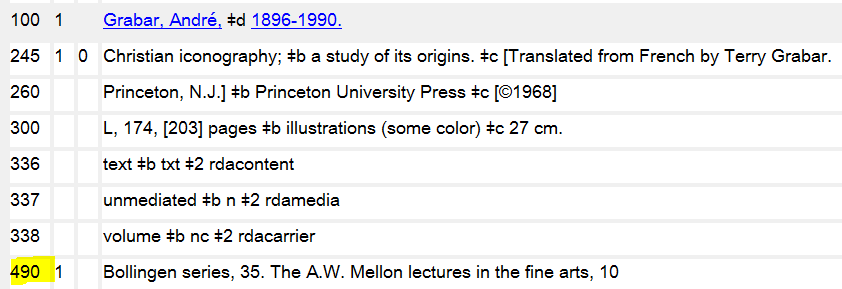
**Its OCLC record (OCLC#: 69)**



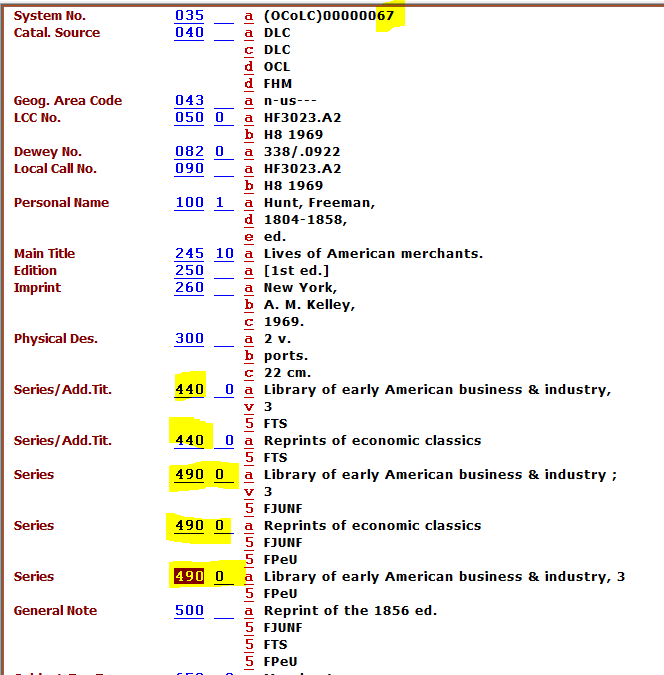
**Example 2: Multiple 490s/830s (same tracing practice) on Aleph Record**



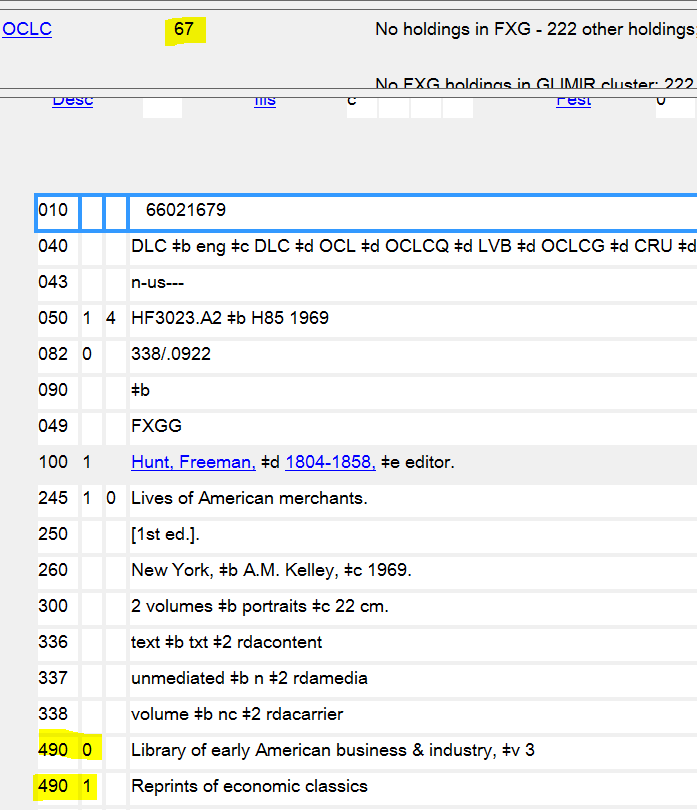
**Its OCLC record (OCLC#: 311)**



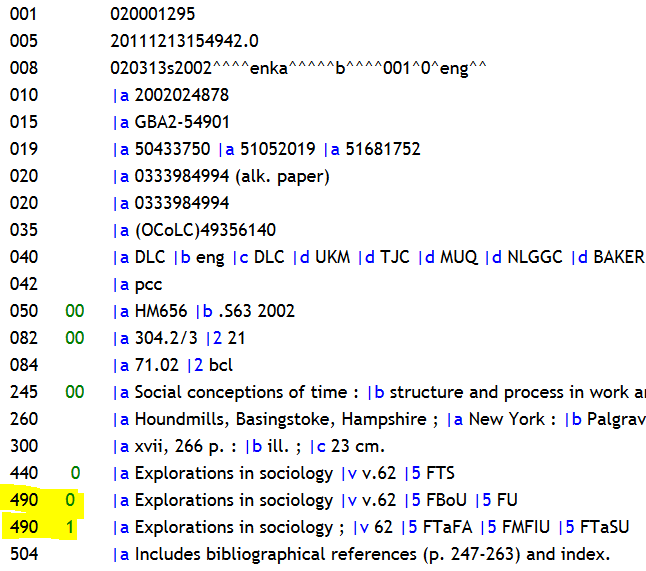
**Example 3: Multiple 490s/830s on Aleph Record**



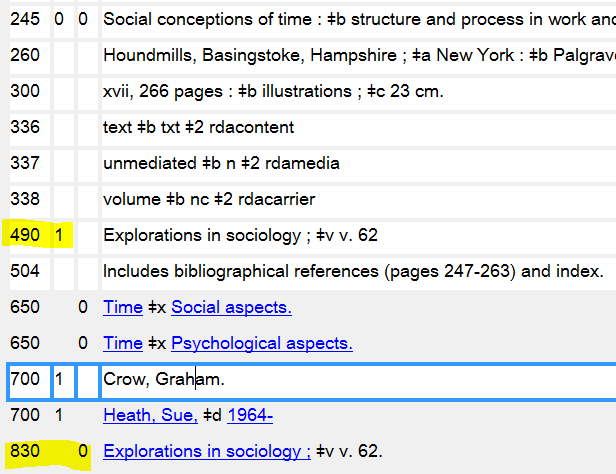
**Its OCLC record (OCLC#: 67)**



**Example 4: Multiple 490s (different tracing practice) Aleph Record**



**Its OCLC record (OCLC#: 49356140)**



1. See https://github.com/EthanDF/FLVC\_490\_Duplicates [↑](#footnote-ref-1)
2. Where xx is the library’s OWN code. This method is still in development by FLVC as of July 8, 2015. [↑](#footnote-ref-2)