

Shared Bib Records with Multiple Series Fields: Issues and Suggested Solution

Prepared by the Bibliographic Control and Discovery Subcommittee (BCDS),
Multiple-Series Cleanup Task Force: Elaine Dong (FIU), chair, Ethan Fenichel (FAU),
Annie Glerum (FSU). Last update, Nov. 18, 2015.

Background

During the merge of the bib records from separate databases of 11 State University Libraries (SULs) in 2012, multiple copies of a record were merged into one record. In order to protect fields that potentially contain unique information, variant text of many fields were kept by adding \$5 with institution codes. As a result, many bib records in the Shared Bib contain multiple series fields (440/490/830) with \$5. These fields are not necessarily local series. A report was run and 209,671 records with multiple series fields (440s and 490s) were identified.

Problems with These Records

- Obsolete 440 field(s) on records (see example 1 below)
- Multiple 490s/830s with same tracing: slightly different statements in the fields were all kept during the Shared Bib merge (see example 2 below)
- Multiple 490 fields with different tracing: due to different tracing practice of SULs (see example 3 below)

Example 1: Obsolete 440 fields on Aleph Record

```
440 0 |a Reprints of economic classics |5 FMFIU |5 FU
440 4 |a The Adam Smith library |5 FMFIU
490 0 |a Reprints of economic classics |5 FJUNF |5 FTaFA |5 FPeU
490 0 |a The Adam Smith library |5 FJUNF |5 FTaFA |5 FPeU |5 FU
```

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

```
490 1 |a Bollingen series, 35:10. The A. W. Mellon lectures in the fine arts |5 FTaSU
490 1 |a Bollingen series, 35. The A. W. Mellon lectures in the fine arts, 10 |5 FSsNC |5 FMFIU |5 FJUNF |5 FPeU |5 FBoU |5 FTaFA |5 FTS |5 FU
490 1 |a Bollingen series, 35:10 |5 FOFT
490 1 |a Bollingen series, 35. The A. W. Mellon lectures in the fine arts, |v 10 |5 FFmFGC
504 |a Bibliography: p. 149-158 (2d group) "Illustrations": p. [1]-[203] (3d group)
505 0 |a The first steps. -- The assimilation of contemporary imagery. -- The portrait. -- The historical scene. -- Dogmas expressed in a single ima
650 0 |a Art, Early Christian
650 0 |a Christian art and symbolism
830 0 |a Bollingen series, |v 35.
830 0 |a A.W. Mellon lectures in the fine arts ; |v 10.
830 0 |a A.W. Mellon lectures in the fine arts.
830 4 |a The A. W. Mellon lectures in the fine arts ; |v 1961
830 4 |a The A. W. Mellon lectures in the fine arts, |v 10
```

Its OCLC record (OCLC#: 311)

490	1	Bollingen series, 35. The A.W. Mellon lectures in the fine arts, 10
776	0 8	#i Online version: #a Grabar, André, 1896-1990. #t Christian iconography. #d Princeton, N
830	0	Bollingen series ; #v 35.
830	0	A.W. Mellon lectures in the fine arts ; #v 10.

Example 3: Multiple 490s (different tracing practice) Aleph Record

440	0	a Explorations in sociology v v.62 5 FTS
490	0	a Explorations in sociology v v.62 5 FBoU 5 FU
490	1	a Explorations in sociology ; v 62 5 FTaFA 5 FMFIU 5 FTaSU
504		a Includes bibliographical references (p. 247-263) and index.

Its OCLC record (OCLC#: 49356140)

490	1	Explorations in sociology ; #v v. 62
830	0	Explorations in sociology ; #v v. 62.

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

Instead of updating these records manually, a more efficient approach would be batch overlaying a selected subset of records with OCLC master records to remove obsolete and redundant series fields. We aim to use automated means to flag records that shouldn't be overlaid and update the records that's fit to overlay.

Sample Record Analysis

We did the following analysis in order to identify what records are fit to overlay and how to identify and separate them:

Step 1 Aleph Record 035 Field Analysis

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

1. Records with 035 OCLC# only (674 records, 67%)
2. Records with both OCLC# and vendor# in 035 field (63 records, 6%): the majority of these records are identified as vendor records. FLVC is currently working on cleaning up this type of records.
3. CIS Microfiche Records (have both 035 OCLC# and 035 gp#, some OCLC# have "x" on the end; 36 records, 3%): these Aleph records (microfiche format) were cloned from OCLC records 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 overlaying method since they don't have OCLC records.

After discussion, we agreed that records in Group 2-4 can't or shouldn't be overlaid by their OCLC master records.

Step 2 Aleph Record Format Analysis

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 developed Python script¹ to 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 See https://github.com/EthanDF/FLVC_490_Duplicates

1. Print format: Safe to be overlaid
2. Electronic format: should be excluded for the following reasons:
 - 1) Sometimes vendor records received from vendor and then loaded into Aleph are different from their OCLC master records;
 - 2) 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)

830	0	Pubblicazioni della Stazione Zoologica di Napoli (1980)
850		DLC
856	4	#u http://firstsearch.oclc.org #z Address for accessing the journal using authorization num
856	4	#u http://firstsearch.oclc.org/journal=0173-9565;screen=info;ECOIP #z Address for acces
856	4 0	#u http://www.blackwell-synergy.com/member/institutions/issuelist.asp?journal=mae
856	4 0	#u http://onlinelibrary.wiley.com/journal/10.1111/(ISSN)1439-0485/issues

Aleph record (sys# 020001910)

830	0	a	Pubblicazioni della Stazione Zoologica di Napoli (1980)
856	4	u	http://firstsearch.oclc.org/journal=0173-9565;screen=info;ECOIP
		u	Address for accessing the journal from an authorized IP address through OCLC FirstSearch Electronic FTaSU
856	4	u	http://firstsearch.oclc.org
		u	Address for accessing the journal using authorization number and password through OCLC FirstSearch FTaSU
856	40	y	Click here for FULL TEXT options
		u	http://sfx.fcla.edu/ucf?sid=sfx%3Ae%5Fcollection&issn=0173-9565
		u	FOFT
856	40	u	http://www.blackwell-synergy.com/member/institutions/issuelist.asp?journal=mae
		u	FTaSU
856	40	u	http://ezproxy.fau.edu/login?url=http://onlinelibrary.wiley.com/journal/10.1111/(ISSN)1439-0485
		u	FBoU

3. 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 script, a total of 17 records found a mismatched form between their OCLC and Aleph records.
4. Unknown format: most of them are map or GIS records. They are safe to be overlaid.

Step 3 OCLC Master Record 490/830 Field Analysis

Among the OCLC master records of the Group 1 records (7,535 records with OCLC# only), we separated the following OCLC records:

1. 83 OCLC records (1%) don't contain any 490/830 fields

Since their Aleph records contain 490/830 fields that might be local series, we agreed that these records are not fit for overlay and should be sent for authority review.

2. 1,222 OCLC records (16%) contain 490 0# fields.

We discovered that many of these 490 0# field already have established series headings, it could be that original catalogers used 490 0# due to their untracing practice regardless the existence of series headings.

After discussion, we decided that these 1,222 (16%) should be excluded from overlay and parsed out for authority review.

Information on Local Series

After collecting information from SUL representatives and colleagues, we came to understand how and what types of local series have been created. There are the following three types of local series created by SULs:

1. Local made-up series: usually created for special collection materials and contain the word "Collection." They don't exist in SUL local shared authority database AUT UXU01

Fields used by 5 SULs (UF, USF, FSU, UCF, FIU): 500, 590, 710, 797, 490 1#/830, and 899. See the following screenshot for an example. Among these fields, 710, 490/830 are not currently protected from OCLC overlay. We didn't find local made-up series added by the other 7 SULs (FAMU, FAU, FGCU, FPU, NCF, UNF, and UWF).

```
=440 \0$aFlorida. Geological Survey. Geological bulletin$vno. 33$5FMFIU$5FFmFGC
=440 \0$aGeological bulletin (Tallahassee, Fla.) ;$vno. 33$5FU
=490 1\1$aFlorida. Geological Survey. Geological bulletin no. 33$5FOFT$5FTaSU$5FPeU
=490 1\1$aGeological bulletin no. 33$5FSTS
=490 1\1$aGeological bulletin ;$vno. 33$5FJUNF
=500 \1$aIncludes index.$5FOFT$5FJUNF
=500 \1$aMaps in pocket.$5FJUNF
=504 \1$aBibliography: p. 247-252.
=530 \1$aElectronic reproduction of copy from George A. Smathers Libraries, University of I
=650 \0$aGeology$zFlorida$zCitrus County.
=650 \0$aGeology$zFlorida$zLevy County.
=710 2\2$aFloridiana Collection.$5FSTS
```

2. Locally created Series before SULs have authority or funnel to establish LC Series headings. These series were put in 440/490/830 fields on Aleph records

Before the Shared Bib merge, some SULs have their own authority database. Many local series were created before SULs have authority or funnel to establish LC series headings. After the merge, all the locally created authority records were migrated to the AUT UXU10 database, which contains both series created by SULs and imported LC series headings. We examined some records in AUT UXU10, and found out that many of them already have established LC series headings, which may have the same or different forms, same or slightly different treatment (e.g. class together or separately). Here is an example for different form:

Locally created series (130) by FIU in AUT UXU10:

<u>040</u>	<u>—</u>	<u>a</u>	FXG
		<u>c</u>	FXG
<u>090</u>	<u>—</u>	<u>a</u>	F1401
		<u>b</u>	.P153
		<u>d</u>	items which are included in Publicación (Pan American Institute of Geography a
		<u>s</u>	FMFIU
<u>130</u>	<u>0</u>	<u>a</u>	Publicación (Pan American Institute of Geography and History).
		<u>p</u>	Instituto Panamericano de Geografía e Historia, Comisión de Historia (Series).
		<u>p</u>	Misiones americanas en los archivos europeos

It was put in 490 1#/830 in Aleph bib #: 025262868

<u>490</u>	<u>1</u>	<u>a</u>	Misiones americanas en los archivos europeos,
		<u>v</u>	12
<u>830</u>	<u>0</u>	<u>a</u>	Publicación (Pan American Institute of Geography and History).
		<u>p</u>	Instituto Panamericano de Geografía e Historia, Comisión de Historia (Series).
		<u>p</u>	Misiones americanas en los archivos europeos,
		<u>v</u>	12.

It has established LC Series heading now which has different form, 830 field on OCLC record is different:

<u>490</u>	<u>1</u>	Misiones americanas en los archivos europeos ; #v 12
<u>830</u>	<u>0</u>	Misiones americanas en los archivos europeos ; #v 12.

3. Series that is treated quite differently from the authority record

We learned from our colleagues that there were some series created differently from the authority records. For example: ARN 4709659 “Evergreen black cat book” instructs to use as a quoted note only, FSU added access point for this “quoted note” heading in 490 1#/899 in Aleph as below:

Aleph#: 23891598

490	1	a	An Evergreen black cat book ;
		v	B-385
		5	FTaSU
899	0	a	Evergreen black cat book ;
		v	B-385.
		5	FTaSU

OCLC#: 1401804: OCLC record has 500 quoted note, but no 490/830 fields

338		volume #b nc #2 rdacarrrier
500		"Evergreen black cat book".

Record Cleanup Plan

Based on the findings from sample record analysis and information collected for local series, we came up with an initial cleanup plan. We tested on the 1st 10k records. Ethan adjusted the program script many times after analyzing the test results, and the following is the finalized cleanup plan:

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

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

Step 2. Identify OCLC records to be sent for authority review

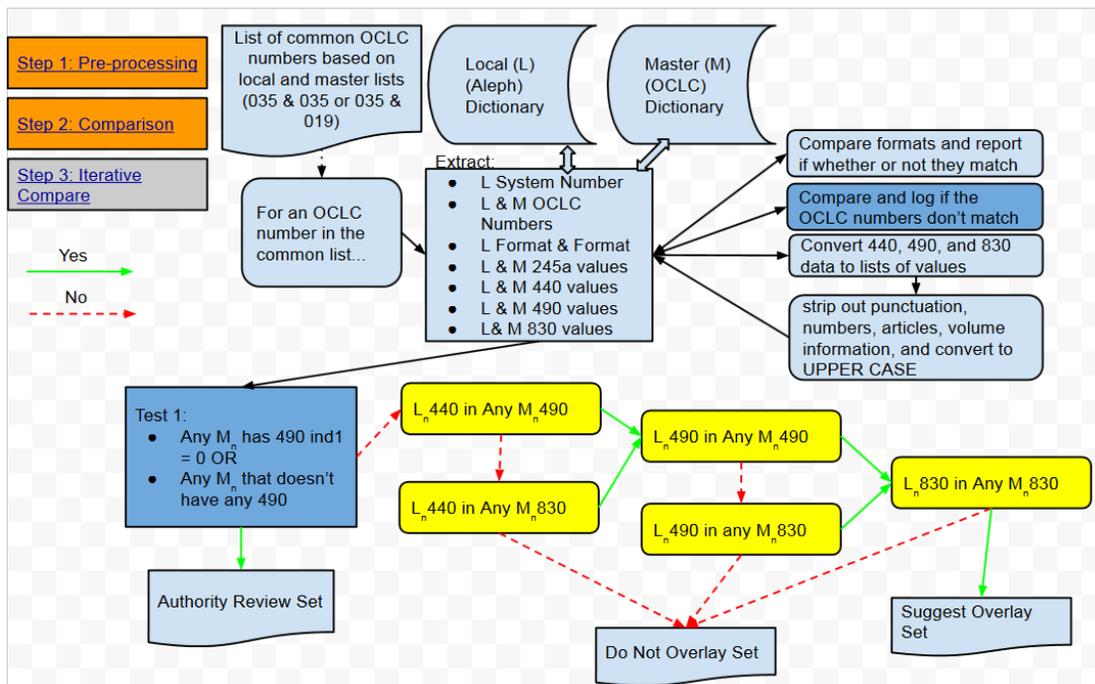
- Run program to identify any OCLC records that contain 490 0# field OR don't contain any 490 field. Their Aleph records will be excluded from overlay.

Step 3. Identify Aleph records with local series to be excluded from overlay

Comparison Logic:

- Compare \$a and \$p of 440, 490, 830 fields only
 - Strip out the following data of \$a and \$p before comparison:
 - The beginning English, French, and Spanish articles: *the, a, an, el, los, la, las, un, unos, una, unas, le, la, l', les, un, une, des*
 - The beginning word: His, Her, Him
 - Most of punctuation marks including ‘ ’ “ ” ... ! ; , . [] < > () { } - | / \
 - All the numbers
 - Convert all text to upper case, then
 - Remove independent words “NO” “V” “VOL”
 - Remove the subfield a from the text string (all cases of “a:”)
 - Normalize diacritical marks
1. Compare all the 440 fields of an Aleph record with 490 and 830 fields of its OCLC master record, if match found, send it/them to Suggest Overlay Set; if no match found, send it/them to Do Not Overlay Set.
 2. Compare all the 490 fields of an Aleph record with 490 and 830 fields of its OCLC master record, if match found, send it/them to Suggest Overlay Set; if no match found, send it/them to Do Not Overlay Set.
 3. Compare all the 830 fields of an Aleph record with 830 fields of its OCLC master record, if match found, send it/them to Suggest Overlay Set; if no match found, send it/them to Do Not Overlay Set.

Diagram for Step 2 & 3



Step 4. Update the 035 OCLC# of Aleph records to be overlaid

- Run program to identify mismatched 035 OCLC# between Aleph and OCLC records. The cause of these mismatched OCLC# was that old OCLC record was merged with another OCLC record. If Aleph have both OCLC records, they should be merged before overlay; if Aleph has only the old OCLC record, the OCLC# on these Aleph records need update so that extracted OCLC records can match on the updated OCLC# of Aleph records and overlay them
- Update Aleph records' OCLC# by Aleph Services function: Global Changes (manage-21) and Load Catalog Records (manage-18)]

Step 5. Create Genload profile for loading MARC data file

The Genload profile sets OCLC# as match point, does not load 9xx and 856 fields from OCLC master records, only overlays the Aleph bib records and not touch holding and item, and protects the following fields:

- 1) Fields with or without \$5: 590, 599, 970, STA, LKR, TKR
- 2) Fields **with \$5**:

| MARC TAG |
|----------|----------|----------|----------|----------|
| 351 | 541 | 690 | 797 | 909 |
| 500 | 545 | 691 | 845 | 951 |
| 501 | 561 | 699 | 856 | |
| 506 | 562 | 710 | 896 | |
| 520 | 563 | 790 | 897 | |
| 533 | 583 | 791 | 898 | |
| 540 | 584 | 796 | 899 | |

- 3) 655_7 Genre heading fields with the following \$2:

MARC TAG	655_7	655_7	655_7	655_7	655_7	655_7	655_7	655_7
\$2	rbprov	rbbin	rbgenr	rbpap	rbpri	rbpub	rbtyp	local

Note: we added several protected fields, including STA and 599 (with or without \$5), 520 \$5 (some SULs have proprietary data in the field), 710 \$5 (some SULs put local series in the field), and 655_7 \$2 local, in addition to the 33 currently protected fields from OCLC overlay as implemented by FLVC.

Step 6. Overlay the subset of records by Genload

- Pull out marc files of OCLC records from the Suggest Overlay Set
- Do the test load first, and then real load in Genload utility to overlay corresponding Aleph bib records

Project Timeline

Jan. 2015	A report of 209,671 shared bib records with multiple series (440/490/830) fields was generated by FLVC
Mid-April	Task Force was formed to solve issues of these records
May—Aug.	Analyzed sample records and collected information
June	Strategy developed: use program to flag records that contain local data, and batch overlay/update the records with obsolete and duplicate series by their OCLC master records
June—Aug.	Developed, tested and finalized the Python program
Last week of Aug.	Configured Genload profile for loading OCLC master records, test loading was successful. Genload profile was approved by FLVC
Sep. 3	New report received from FLVC includes 222,404 shared bib records with multiple series statements
Sep.	Analyzed new report. Ran the Python program, identified the following: <ul style="list-style-type: none"> • 53,802 records as candidates for overlay • 111 duplicate records (035/019 dup) from Overlay Set, sent to Melissa Stinson at FLVC for merge • 1,123 records: OCLC# on Aleph records don't match OCLC master record number due to the merge of OCLC records. Will update their OCLC# before overlay
Oct. 8	Present the project at CAM meeting, ask for review of Overlay Set and 1,123 records with mismatched OCLC# for approval
Oct. 12	Present the project at TSSC meeting and collect feedback
Nov.	Load the OCLC master records from Overlay Set into Aleph to overlay/update problematic records

Cleanup Result

The program identified 53,802 records as candidates for overlay out of the 222,404 records with multiple series fields, after excluding the records that are not fit for overlay. A total of 51,818 bib records in Aleph were overlaid, the obsolete 440 fields and unnecessary multiple 490 and 830 fields on these records got updated.

Here is an example of Aleph record before and after overlay:

Aleph Record before Overlay	Aleph Record after Overlay Note: 440/490/830 were updated; local field such as 899, 951 fields were protected
<pre> 440 0 Readings in Oriental thought FTS FTaSU 440 4 The Modern library of the world's best books <205> FTS 490 0 The Modern library of the world's best books [205] FTaSU FPeU FU FMFIU 490 0 Readings in Oriental thought FPeU FJUNF FBoU FU FMFIU 490 0 The Modern library of the world's best books FBoU 490 1 The Modern library of the world's best books [205] FJUNF 504 Bibliography: p. [399]-401. Bibliographical footnotes. 650 0 Buddhism Collections. 650 0 Buddhism Sacred books. 830 0 Modern library of the world's best books; [205] 830 0 Readings in Oriental thought. 899 0 Wedig collection. FMFIU NCO 5 951 02 FAU01:000255331; FBoU 951 04 FIU01:001349334; FMFIU 951 05 FSU01:000547336; FTaSU </pre>	<pre> 490 1 Readings in Oriental thought 490 1 The Modern library of the world's best books [205] 504 Includes bibliographical references (pages 399-401. l 505 0 Early Buddhism -- The life of Buddha as a way of salv Tantricism and the decline of Buddhism in India -- The -- The introduction of Buddhism to Japan -- Saicho ar the pure land -- Nichiren's faith in the lotus -- Zen. 650 0 Buddhism Sacred books. 830 0 Modern library of the world's best books ; 205. 830 0 Readings in Oriental thought. 899 0 Wedig collection. FMFIU NCO blank 951 02 FAU01:000255331; FBoU 951 04 FIU01:001349334; FMFIU 951 05 FSU01:000547336; FTaSU 951 09 NFU01:000231416; FJUNF 951 10 SFU01:000000770; FTS 951 08 UFU01:000812032; FU 951 11 WFOU01:000222854; FPeU </pre>