CATACODERS

Tech-Savvy Catalogers Remediate Messy Series Fields to Alleviate Display Issues in the State University System of Florida Libraries Consortial Interface
I. Multiple-Series Cleanup Task Force

II. Issues with initial merge into “Shared Bib”

III. Analysis of local and OCLC records

IV. Implementation of batch overlays

V. Advantages of being “catacoders”
TASK FORCE
Multiple-Series Cleanup Task Force

• BCDS ⇔ CAM ⇔ CSUL
• *Guidelines and Procedures for the Shared Bibliographic Catalog*
• Differing tracing practices
• Multiple 4XX fields in shared records
In a library catalog database, quality bibliographic description provides both accurate and sufficient enough data such that the library user can discover the description, identify the resource, and then determine whether that resource fulfills the user’s needs.
ISSUES
Initial Shared Bib merge

- All eleven universities’ libraries sharing one DB
ONE SHARED BIB DB

Items, Orders
However…

*We know all universities are not created the same*
ELEVEN UNIVERSITIES

- Florida Agricultural and Mechanical University
- Florida Atlantic University
- Florida Gulf Coast University
- Florida International University
- Florida State University
- New College of Florida
- University of Central Florida
- University of Florida
- University of North Florida
- University of South Florida
- University of West Florida
ELEVEN STUDENT BODIES

- Florida Agricultural and Mechanical University
- Florida Atlantic University
- Florida Gulf Coast University
- Florida International University
- Florida State University
- New College of Florida
- University of Central Florida
- University of Florida
- University of North Florida
- University of South Florida
- University of West Florida
ELEVEN COLLECTIONS

- Florida Agricultural and Mechanical University
- Florida Atlantic University
- Florida Gulf Coast University
- Florida International University
- Florida State University
- New College of Florida
- University of Central Florida
- University of Florida
- University of North Florida
- University of South Florida
- University of West Florida
Initial Shared Bib merge

- Eleven universities sharing one bib database
- FLVC and CSUL SB Working Group erred on the side of caution to protect local data
  - Used $5 [MARC Org Code] to protect fields and trigger display
  - Resulted in duplicated “near-match” data
ISSUES

In MARC records

- Differing transcriptions
- Incorrect subfield coding
- Incorrect ISBD punctuation
- Use of obsolete 440 fields
- Varying tracing practices
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 $5 FSsNC

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

[...]

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.

440 _0 $a Reprints of economic classics $5 FMFIU $5 FU

440 _4 $a The Adam Smith library $5 FMFIU

490 1_ $a Reprints of economic classics

490 1_ $a The Adam Smith Library

440 _0 $a Explorations in sociology $v 62. $5 FTS

490 _0 $a Explorations in sociology $v 62. $5 FBoU $5 FU

490 _1 $a Explorations in sociology $v 62.
ISSUES

In Mango OPAC Display

- If every **490** has a **$5** , then the field does not display in union Mango
- If both **440** and **830** are present, then both series display in union and local Mango
- If a correct field has another library’s **$5** code, then it doesn’t show in local Mango
Fairs, past and present; a chapter in the history of commerce.


x, 318 p. 22 cm.

Series: Essays in history, economics, & social science, 8.

Selected essays in history, economics, & social science, 8.
Author, etc.: 
Walford, Cornelius 1827-1885

Title: 
Fairs, past and present; a chapter in the history of commerce.

Published: 

Description: 
x, 318 p. 22 cm.

Series: 
Burt Franklin research & source works series no. 163
Essays in history, economics, & social science, 8.
ANALYSIS
FLVC generated a CSV file of records with multiple 4xx fields (209,671 rows)

- Analyzed report & catalog records
- Initial idea was that catalogers would correct each record
- Explored feasibility of automating the decision process and cleanup

MarcEdit
Excel
Python
Comparing local and OCLC Records (master)

- A given record from our catalog with *certain attributes*
- Compare *contents* to the master record and if it met *certain criteria*, we would overlay our local record with the master
- How to download the records from OCLC?
ANALYSIS

• Generated a 1000 record sample set
  • Excel VBA macro \((\text{MOD(ROW(),209)=1})\)

• Pulled sample set of records from catalog
  • Analyzed and categorized records using Excel and MarcEdit
  • Identified potential local series and other cases we wouldn’t want to overlay
Pulled batches of OCLC records with MarcEdit’s Z39.50 utility
Refining the Python script

• Developing with a little help from my friends:
  • *PyMarc Google Group*
    eg: *Diacritics* ⇒ á, é, í, ó, ú, ü, ñ
  • *Twitterverse*

• Diagrammed the decision making process using *Google Drawings*
Decision Tree 1: Preprocessing
Decision Tree 2: Comparison

- OCLC_Numbers.csv
  - Filter out OCLC numbers associated with certain formats
  - Pass list to MarcEdit Z39.50 protocol to download OCLC MARC Records

- Corresponding Marc records from Aleph for the OCLC_Numbers.csv
  - Return Marc records from OCLC (some will be the merged records from the provided OCLC Number)

- Read into 2 Dictionaries & Lists for local (Aleph) and master (OCLC) records
  - Produce list of common OCLC numbers based on local and master lists (035 & 035 or 035 & 019)
List of common OCLC numbers based on local and master lists (035 & 035 or 035 & 019)

For an OCLC number in the common list...

Extract:
- L System Number
- L & M OCLC Numbers
- L Format & Format
- L & M 245a values
- L & M 440 values
- L & M 490 values
- L & M 830 values

Compare formats and report if whether or not they match

Compare and log if the OCLC numbers don't match

Convert 440, 490, and 830 data to lists of values

Strip out punctuation, numbers, articles, volume information, and convert to UPPER CASE

Test 1:
- Any M<sub>n</sub> has 490 ind1 = 0 OR
- Any M<sub>n</sub> that doesn't have any 490

Authority Review Set

L<sub>n</sub> 440 in Any M<sub>n</sub> 490

L<sub>n</sub> 440 in Any M<sub>n</sub> 830

L<sub>n</sub> 490 in Any M<sub>n</sub> 490

L<sub>n</sub> 490 in Any M<sub>n</sub> 830

L<sub>n</sub> 830 in Any M<sub>n</sub> 830

Yes

Do Not Overlay Set

No

Suggest Overlay Set

Decision Tree 3: Iterative compare
ANALYSIS

https://github.com/EthanDF/FLVC_490_Duplicates
ANALYSIS

Testing the Python report

• Ran test of 1000 records
• Refined the script
• Confirmed automated process was feasible
• Ran test of 10,000 records
  • Ruled out additional changes
  • Tested computer's processing limit
IMPLEMENTATION
IMPLEMENTATION

Preparing for implementation

- Presented project update to BCDS
- Pulled 23 batches of MARC catalog records
- Refined the workflow and scripts
- Configured batchloading profile
  - Selected fields to keep
  - Got approval by FLVC
# IMPLEMENTATION

## Fields to keep in the catalog record

<table>
<thead>
<tr>
<th>MARC TAG</th>
<th>MARC TAG</th>
<th>MARC TAG</th>
<th>MARC TAG</th>
<th>MARC TAG</th>
<th>MARC TAG</th>
</tr>
</thead>
<tbody>
<tr>
<td>351</td>
<td>541</td>
<td>690</td>
<td>797</td>
<td>909</td>
<td></td>
</tr>
<tr>
<td>500</td>
<td>545</td>
<td>691</td>
<td>845</td>
<td>951</td>
<td></td>
</tr>
<tr>
<td>501</td>
<td>561</td>
<td>699</td>
<td>856</td>
<td></td>
<td></td>
</tr>
<tr>
<td>506</td>
<td>562</td>
<td>710</td>
<td>896</td>
<td></td>
<td></td>
</tr>
<tr>
<td>520</td>
<td>563</td>
<td>790</td>
<td>897</td>
<td></td>
<td></td>
</tr>
<tr>
<td>533</td>
<td>583</td>
<td>791</td>
<td>898</td>
<td></td>
<td></td>
</tr>
<tr>
<td>540</td>
<td>584</td>
<td>796</td>
<td>899</td>
<td></td>
<td></td>
</tr>
<tr>
<td>655_7 $2</td>
<td>rbprov</td>
<td>rbbin</td>
<td>rbgenr</td>
<td>rbpap</td>
<td>rbpri</td>
</tr>
</tbody>
</table>
Following a feedback and review period…

Within two weeks 51,818 local records were overlaid with OCLC master records
CATACODERS
Being a “tech-savvy” cataloger…

- Allows for independence in developing custom solutions on your time schedule
- Facilitates communications with computer programmers and other technology staff
- Develops essential skills for library linked data environment and future discovery tools
“The code4lib community talks about getting catalogers and coders together, and wonders if catalogers should be taught to code. The problem, however, is not between coders and catalogers but is systemic in our profession.

We have kept cataloging and computer technology separate, as if they aren't both absolutely necessary. One is the chassis, the other the engine, but nothing useful can come off the assembly line unless both are present in the design and the creation of the product.”

–Karen Coyle

http://kcoyle.blogspot.com/2015/05/catalogers-and-coders.html
RESOURCES
SOURCES

Project documentation. https://sharedbib.pubwiki.fcla.edu/wiki/index.php/Cataloging#Multiple_Series_Fields_Cleanup

Project Github. https://github.com/EthanDF/FLVC_490_Duplicates

All about Genload. https://fclaweb.fcla.edu/node/521


PyMARC Google Group. https://groups.google.com/forum/#!forum/pymarc


Glerum, Annie. *Adding XSLT to the Cataloger’s Toolbox*. http://connect.ala.org/files/Adding%20XSLT%20to%20the%20cataloger%27s%20toolbox-Final.pdf


ACRONYMS

BCDS = Bibliographic Control and Discovery Subcommittee

CAM = Cataloging, Authorities, and Metadata Committee

CSV = Comma Separated Values

CSUL = Council of State University Libraries [of Florida]

DB = database

FCLA = Florida Library Automation Center

FLVC = Florida Virtual Campus

ISBD = International Standard of Bibliographic Description

OCLC = Online Computing Library Center

MARC = MAchine-Readable Cataloging

SB = Shared Bib [i.e. The Shared Bib Catalog]
MULTIPLE-SERIES CLEANUP TASK FORCE

Elaine Dong, chair
edong@fiu.edu

Ethan Fenichel
fenichele@FAU.EDU
@EthanDF

Annie Glerum
aglerum@fsu.edu
@annie_glerum

State University System Libraries of Florida

http://union.discover.flvc.org/