October, 2018

Reclassing Dewey to LC

Jennifer Eustis, *University of Massachusetts Amherst*

bepress™

Reclassing Dewey To LC

That Time When Alma Was Really Your Best Friend

Jennifer Eustis
jeustis@umass.edu

Outline

- Introduction & Background
- The Puzzle Pieces
    - Information Update Job
    - Alma Analytics
    - OCLC's API
- Assembling the Puzzle
- Outcomes
- Statistics
- What we've learned
- Questions

Dewey Reclassification Project

"Image from page 814 of 'North Carolina Christian advocate [serial]' (1984)"

https://www.flickr.com/photos/internetarchivebookimages/14781527524

Dewey Reclassification Project

Project Homepage:
https://confluence.uits.uconn.edu/confluence/display/LCD/Reclassification+of+Dewey+Class+items

The Problem

"The Master Plan Astrop-Buescher State Park"

https://www.flickr.com/photos/texasstatearchives/6145312610

We wanted to find a way to reclass a significantly large number of Dewey items to LC. We wanted to rely when possible on automated processes.

- The UConn Library Master Plan
- EAST
- Collections Review
- Duplicates
- Unusable books still in circulation
- Space
- Space …

The Dewey collection was a collection apart from the main circulating stack. In short, we have 2 circulating stack collections.

EAST: Our Pilot for the Dewey Reclassification

"Image from page 127 of 'The world almanac and encyclopedia (1918)'"

https://www.flickr.com/photos/internetarchivebookimages/14579562069

Eastern Academic Scholars Trust

"...a shared print initiative across 60 partner libraries, primarily in the Eastern United States. EAST directly addresses the growing need for academic and research libraries to ensure that print monographs and journals of scholarly value are not inadvertently discarded as libraries undertake necessary weeding and deselection programs to free up space for other library services." - About page

EAST Project Homepage:
https://confluence.uits.uconn.edu/confluence/display/LCD/EAST+LC+Classed+Dewey+Project

5

Goal of EAST Pilot

"#0a Production Flight Test, Best Hart, Mort, McNeil, Grier, Worley"

https://www.flickr.com/photos/sdasmarchives/9135562609

Pilot Project:

- Identify trends
- Create a benchmark
- Work to make processes more efficient

Our assumptions about reclassifying the Dewey collection:

- Bibliographic record quality often poor
- Items not associated with correct record
- Missing items
- Unusable items
- Master record in WorldCat would have a LC call number and be of better quality

6

Figuring Out The Pieces

Information Update Job: Puzzle Piece No. 1

		Name	Description	Content Type	Type
1	○	Rebuild item description	Rebuild item description	Physical item	Information Update
2	○	Move electronic portfolio information	Move electronic portfolio information	Electronic collections	Information Update
3	○	Global Representation Changes	Global representation changes for bib set	Digital title	Information Update
4	○	Create physical item move requests	Update item information and initiate moving a set of physical items to a new location.	Physical item	Information Update
5	○	Change physical items	Update item information for a set of physical items.	Physical item	Information Update
6	○	Change Holding Information	Update holding information for a set of physical items	Physical item	Information Update
7	○	Change electronic portfolio information	Update portfolio information for a set of electronic portfolios	Portfolio	Information Update

Type : **Information Update**   Source type : **All**

Change Holding Information

Task Parameters: Change Holding Information	⌄

Apply only for Library	[_____ ▾]	location	[- ▾]

☐	Update call number from Bibliographic record	

☐	Correct the data using normalization rules	[_____ ▾]
☐	Tag records for publishing	[**Suppress** ▾]

Pros and Cons to 1st Puzzle Piece

Pros:

- Ability to update holding record with LC call number found in bibliographic record
- Ability to correct data using normalization rules
- Ability to suppress holdings records

Cons:

- Unable to suppress bibliographic record if that record has only 1 holdings record
- Needed a managed set of items
- Unable to run more than one option at a time
- Issues when multiple call numbers were present in bibliographic record

Requirements for Alma's Change Holding Information

- Needed a set of items
  - Fit our requirements of being retained to EAST and part of the Dewey collection
- Had to run the job multiple times if we wanted to enrich the records using a normalization rule and suppress the holdings
- Had to find a way to suppress those bibliographic records with only 1 holding record
- Had to ensure that everything was coded for LC call numbers in both bibliographic record and holding record
- Ensure that only 1 call number was present in bibliographic record
- Knowledge of how to run Alma jobs, normalization rules, processes

Alma Analytics: Puzzle Piece No. 2

We needed a set of active items that fit both being retained for EAST and Dewey. Further, we needed a list of items with barcode information, title, author, MMS ID, OCLC number. Why?

- This list needed to be used for the various puzzle pieces...
    - In Alma to create an item managed set to run the change holding information jobs on
    - For our students who were getting items from the stacks
    - For our end processing unit
    - Retrieving master records from WorldCat

We needed a lot of information that a managed set in Alma wasn't able to provide.

Alma Analyses

Pros and Cons to 2nd Puzzle Piece

Pros:

- Ability to create a list of items that fit our requirements

Cons:

- Lists revealed a number of issues
  - Call number issues (LAW, Law, Microfiche, Unknown, UNKNOWN,...)
  - Lifecycle
- Could we find a way to suppress a bibliographic record that had only 1 holding record that was suppressed?

Requirements for Alma's Analytics Analyses

- Needed to do some initial cleanup
  - Remove bad information in 050 field
  - Check on what was designated as Microfiche/microfilm and Law
  - What did unknown mean?
- Remember to select only active bibliographic records and items
- Knowledge of Alma Analytics

WorldCat API: Puzzle Piece No. 3

"Burglars Wanted"

https://www.flickr.com/photos/nsarchives/7976733112

We needed to get LC call numbers. Our assumption was that the WorldCat master record had them. But we had to be careful of any call number.

A call number:

- What type of call number is needed?
- Where is that information stored in the bibliographic record?
- Where is that information stored in the holding record?
- How is that information encoded in both the bibliographic and holding record?

16

Getting the LC Call Number

- Seeking information in the 050
- Where that LC call number came from was not important (the indicators)

The problem was/is that these are older records. If there was a call number in the record, it was either a Dewey assigned number or one assigned by the cataloger at that time.

Many of the records had/have multiple bibliographic marc call number fields (050, 051, 082, 083, 086, 099)

050-099 Introduction to Call Numbers

Record Type
BIB

Call number fields This section explains the use of call numbers in the following call number fields:

050	Library of Congress Call Number (R)
055	Classification Numbers Assigned in Canada (R)
060	National Library of Medicine Call Number (R)
070	National Agricultural Library Call Number (R)
080	Universal Decimal Classification Number (R)
086	Government Document Classification Number (R)
090	Locally Assigned LC-type Call Number (R)
092	Locally Assigned Dewey Call Number (R)
096	Locally Assigned NLM-type Call Number (R)
098	Other Classification Schemes (R)
099	Local Free-Text Call Number (R)

https://www.oclc.org/bibformats/en/0xx/050-099.html

17

Getting LC Call Numbers - OCLC's APIs

WorldCat Metadata API

The WorldCat Metadata API is a read-write service for WorldCat that makes it possible to add or update master bibliographic records in WorldCat, maintain holdings information, and work with local bibliographic data.

Status: Production   **Sandbox access:** Yes

This service allows libraries to write MARCXML bibliographic records to WorldCat, either as new records or updated records that are replacing existing records. Existing WorldCat bibliographic records can be accessed with a read operation. The API enables an institution to set and delete its holdings on WorldCat bibliographic records. Local bibliographic data

Try the API Explorer >>
Interact with this API

Request a key >>
Request and manage your WSKeys

https://www.oclc.org/developer/develop/web-services/worldcat-metadata-api.en.html

Our records for this pilot project and even for the larger Dewey collection have OCLC numbers. The state of our records were questionable and very old - dating to the 1940-1960's. We bet that their master records since that era had been enhanced. So we needed:

- WorldCat Metadata API Key
- List of OCLC numbers
- Download master records
- Enrich any records without a LC call number
- Overlay these records on our records in Alma

18

MarcEdit to the Rescue

https://marcedit.reeset.net/oclc-integration-information

MarcEdit -- Your complete *free* MARC editing utility

Home
Downloads
Features
Getting Help
  › Help
  › Knowledge Base
  › MarcEdit 101 Webinar
  Series
  › Tutorials
MarcEdit FAQs
Technical Details
About
  › About MarcEdit
  › About the Author
  › MarcEdit 7 User Map
  › Privacy Policy
Contact Me

Search …

OCLC Integration Information

MarcEdit's OCLC integration allows users to work directly against OCLC's WorldCat database to:

- Set and unset holdings
- Create and Update Master Bibliographic records
- Create, Update, and Delete Local Bibliographic records.

To use the integration option, users need to request that their OCLC Key's support the following API:

- Registry API — MarcEdit uses the Registry API to handle token creation and resolve the Registry API and OCLC Symbol as appropriate. This is new as of July 1, 2018 due to OCLC putting this API behind their permission system.
- Metadata API — MarcEdit uses the Metadata API to handle editing and creation of holdings and bibliographic records
- Search API — MarcEdit uses the Search API to search for records in WorldCat.

These three API are required in order for MarcEdit's OCLC Integration to function correctly. As of Sept. 1, 2018, the OCLC Integration will now check access to all three of these API prior to "validating" the function.

MarcEdit to the Rescue ... Again

https://www.oclc.org/developer/gallery/marcedit.en.html

MarcEdit

MarcEdit is a free metadata editing suite that provides integration with a number of OC

Currently, the program makes use of OCLC's Classify and FAST web services to provide automatic batch classification and description of bibliographic records. Additionally, MarcEdit makes use of the WorldShare Platforms WorldCat Search API and Metadata API to provide batch support for holdings management and bibliographic record editing of data found in WorldCat directly from within MarcEdit.
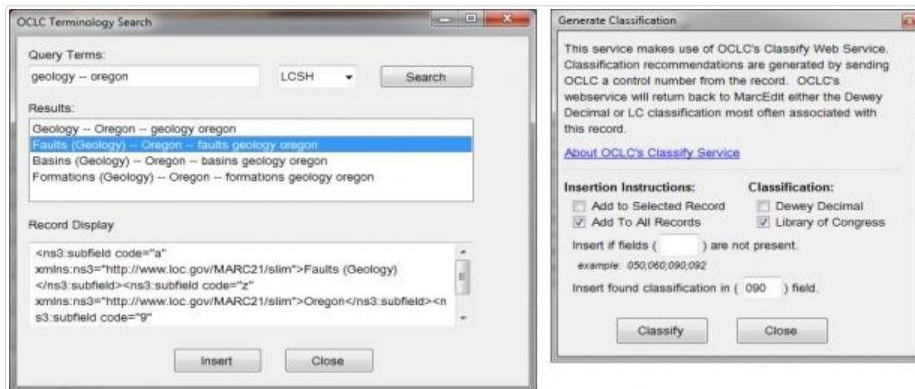
And Again To the Rescue
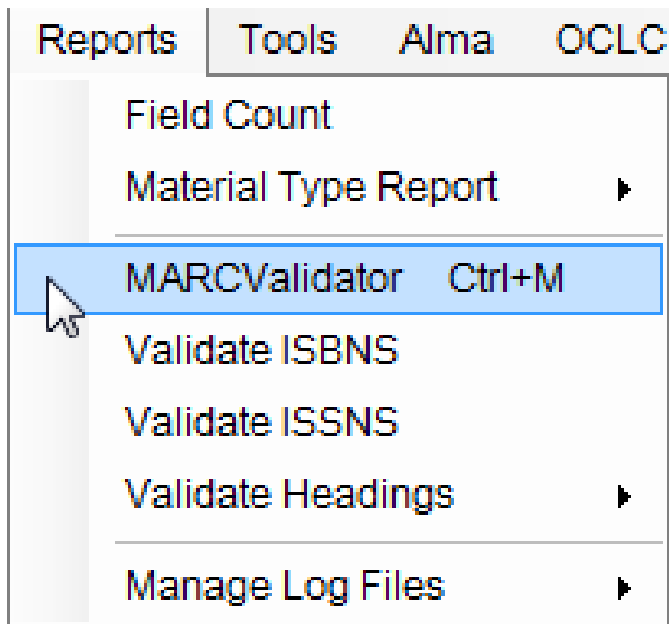
Reports were crucial. Used:

- Field Count
- MarcValidator

These tools helped ascertain a profile of the batch of downloaded master records and whether or not they were valid!

Pros and Cons to 3rd Puzzle Piece

Pros:

- MarcEdit's ability to use
    - OCLC's metadata and classify API
    - Reports
    - Validate Marc Records

Cons:

- Complex process with multiple steps
- No recent documentation on how to set up MarcEdit
- Using the API often timed out

Requirements for WorldCat API

- Needed a text file of OCLC numbers
  - In our case, we had 15 batches or 15 text files that became 15 Marc (mrc) files
- Patience
  - API often timed out
- Add missing LC call numbers using OCLC Classify
- Enrich records with FAST headings
- Validate records using MarcEdit's validation tools
- Extract problem records using MarcEdit's select and extract tools
- Devise an analysis to find those bibliographic records with only 1 holding record
- Knowledge of MarcEdit and how to set up APIs in MarcEdit

The Process: Assembling the Pieces

- Initial Prep cleaning
- Create spreadsheet of items using Alma Analytics
  - Fits project requires and can be used by metadata, end processing, and people searching for items
- Extract OCLC number from spreadsheet
- Download master records and enrich them
- Import and overlay old records with enriched records
- Create item managed set in Alma
- Run update LC call number from bibliographic records on set
- Suppress the holdings record
- Suppress those bibliographic records with only 1 holding record
- Start relabelling to then unsuppress items to send back to regular stacks

Phew! Time for a break!

Lithographed by F. A. Brockhaus, Leipzig (Germany).

THE SPECTRE AND CIRCULAR RAIN BOW.

"Image taken from page 227 of the 'In the Forbidden Land…Second Impression'"

https://www.flickr.com/photos/britishlibrary/11245060544

Initial Outcomes

Goal of the Pilot Project:

- Identify trends
    - Percentage of master records with LC call numbers
    - Number of enriched records
    - Success rate of OCLC's APIs (Classify and Metadata)
- Create a benchmark based on batches 1-10
    - For both the automated piece
    - For end-processing
- Work to make processes more efficient
    - Identified snags along the way

Statistics for End Processing

- Hired two additional work study students for a total of 4 students
- Got additional relabel machines working
- Average to relabel 500-700 items per week
- Approximate each batch to be relabelled had .14% errors.
- Over 26+K items to process and relabel which averages out to about a 8-9 month project
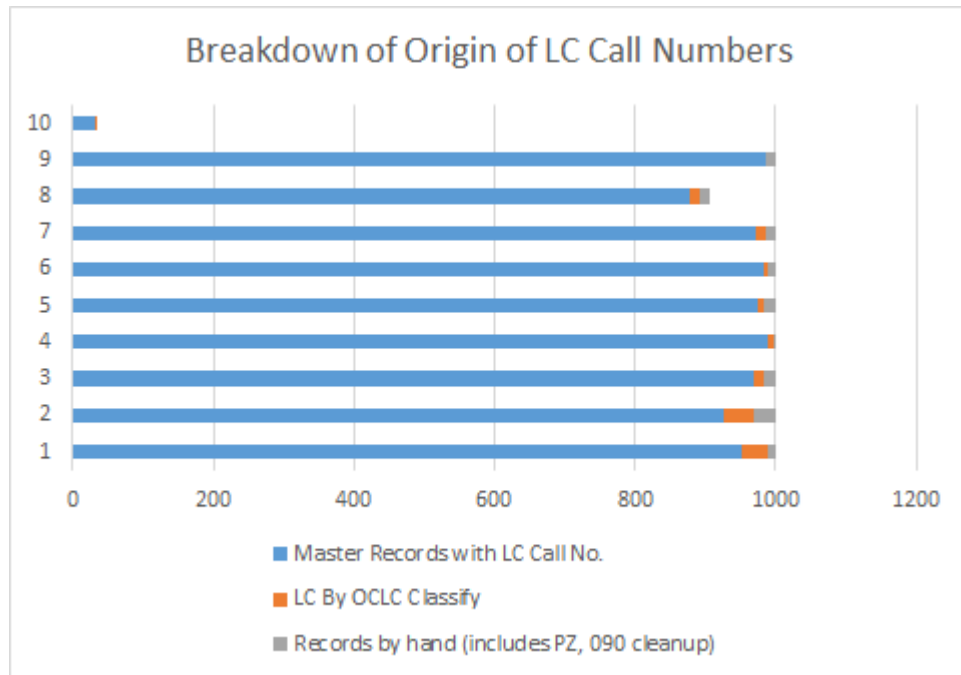
Statistics for OCLC APIs

Success Percentages:

- Average of found LC Call Number: 98.49%
- Average LC call numbers done by hand: 2.7%

Timeout Issues:

- Timeout error occurred on any batch independent of size of text file (500, 1000, 2000, 3000, ... OCLC numbers) or time of day (morning, afternoon, evening)

Breakdown of Origin of LC Call Numbers

■ Master Records with LC Call No.
■ LC By OCLC Classify
■ Records by hand (includes PZ, 090 cleanup)

Number of LC Call Numbers Found

There was one file where the number was unusually low. This proved to be the major exception.

Here, the numbers show that most fall within 98 to 99% of records where a LC call number was found either through OCLC Classify or in the master record downloaded.

Percentage of LC Numbers Found (Master and Classify)
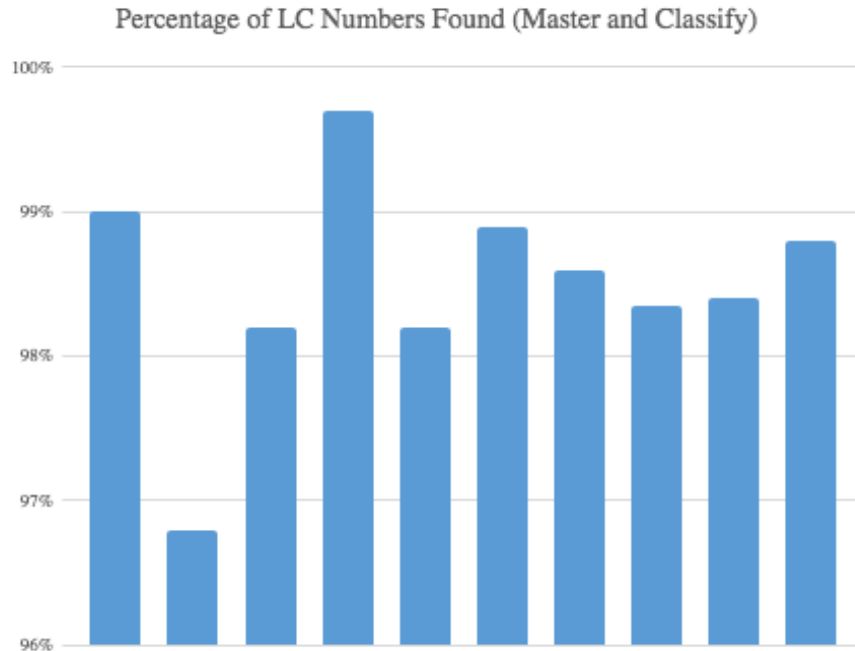
Statistics for Import

An import profile was used to overlay records. The match used was the MMS ID.

There were very few errors in relation to the total number of records overlaid.

What were the errors:

- Electronic and print record mixup
- Incorrect OCLC holdings
- Holdings spread on multiple duplicate bibliographic records

Number of Total Imports and Errors

64

19906

Trends

- Identify trends
  - Percentage of master records with LC call numbers
  - Number of enriched records
  - Success rate of OCLC's APIs (Classify and Metadata)

Our original assumption that the OCLC master record had been enriched and had a LC call number since our records were created was correct. Overwhelmingly, the master records had LC call numbers, summaries, and sometimes added access points. Thanks to OCLC's Classify, finding LC call numbers and record enrichment was a success.

Benchmarks

- Create a benchmark based on batches 1-10
  - For both the automated piece
  - For end-processing

Based on the trends, we now created a benchmark. Approximately, 93-96% of OCLC master records will have LC call numbers. Using OCLC Classify to enrich these records brings that percentage to 98%.

This benchmark was seen in action for batches 11-15 which closely followed this success rate.

End processing was steady as long as our students' work hours were steady. Students were able to resolve many of the quirky LC call number errors.

Efficiences

- Work to make processes more efficient
    - Identified snags along the way

Many of the issues and resulting efficiencies were seen in batches 1 and 2.

- Update LC Call number job works best when only 1 LC call number appears and when all other call number fields are absent. This was addressed in MarcEdit to remove all other call number fields.
- Found that documentation is out of date and hard to follow. We contacted OCLC and posted on the MarcEdit listserv for help. It's OK to feel like an idiot. Created personal reminders about steps and what to watch out for.
- Enriched the records in MarcEdit as the bulk of the work took place there.
- For next project dealing with the rest of the Dewey collection, suggested reviewing the collection first before doing this project.

What did we learn?

- Alma Analytics - creating reports for multiple purposes and re-use
- Documenting practices and its value
- Advantages to splitting up a huge project into smaller pieces and using that first piece as a pilot
- Automation
  - Timeouts with OCLC API
  - Problem of multiple call numbers in bibliographic record
- Lack of automation
  - PZ collection

Questions?

Image from page 255 of

"The Roxburghe ballads" (1879)

Jennifer Eustis
jeustis@umass.edu