

Post-migration Data Clean-up in FOLIO

Lynne Fors and Kara Hart

Wellesley College

WolfCon 2023



Land Acknowledgement

We acknowledge that Wellesley College is built on ancestral and traditional land of the Massachusett people. We also recognize that the United States' removal, termination, and assimilation policies and practices resulted in the forced settlement of Indigenous lands and the attempted erasure of Indigenous cultures and languages.



FOLIO at Wellesley College

- Migration journey began August 2021
 - Legacy system: Millennium
- Implementation team of 7 staff members across the library
 - Access Services, Electronic Resources, Systems, Cataloging & Metadata, Administration
- Go live: June 6, 2022
 - Kiwi release; upgraded in July 2022 to Lotus
 - Went live without Acquisitions apps: Finance, Orders, Invoices
 - Currently implementing Acquisitions functionality
- Clapp Library Renovation announced May 20, 2022
 - Originally projected to begin August 2023 and end July 2024



Our Systems

Pre-Migration

Millennium

Millennium's catalog

Summon

Go Live

FOLIO hosted by EBSCO

VuFind catalog

EBSCO Discovery Service (EDS)

Panorama (for reporting)



Identifying Areas of Clean-up

- We spent *a lot of time* cleaning up our data in Millennium prior to migration
 - standardization, simplification, and plans to leave behind
- During migration
 - 2 test full loads of our data before the final migration
 - Each iteration resulted in clean-up in Millennium to prep for the final migration in May 2022
- These areas were things that could not be done with the migration mappings or we realized we wanted post-migration:
 - Call number types migrated as “Unmapped”
 - Negative implications for browsing and sorting by call number type (future functionality)
 - Holdings record type migrated as “Unmapped”
 - Identifying our EAST retention commitments as a group
 - Cataloged date



Time Saving Tips

- Create a list of one-of-each and unique records
 - Helpful for creating mapping
 - Testing QA of record mapping
 - Testing new user interfaces
- Create a spreadsheet of critical data points with their FOLIO codes, UUIDs, and any notes to document their use or purpose. For example:
 - Locations in FOLIO and your Legacy System
 - Holdings types and Call number types
 - Statistical codes
 - Material types and Note types
 - Patron groups
 - Loan types
 - Institution, Campus, and Libraries
 - Instance and Item Statuses



FOLIO & Millennium Location Mapping

	A	B	C	D	E	F	G	H	I
1	FOLIO SERVICE PT	FOLIO LOCATION CODE	FOLIO LOCATION NAME	FOLIO UUID	FOLIO LIBRARY	FOLIO CAMPUS	FOLIO INSTITUTE	Mill CODE	Mill NAME (call # label too)
2			**A name of a loan policy, a patron group, a material type, a loan type or a location can contain only these characters: a-z, A-Z, 0-9 and dash "-".						
3	Clapp	wcarchives	Archives - 4th Floor Clapp	75940afa-6e3e-4aa0-96d9-47ee3edb6e4e	Archives (archives)	Main Campus (wellic)	Wellesley College (wc)	archv	Archives
4	Art	artresdesk	Art Closed Reserves	497b84ea-a164-47fd-ad6f-228b0bf61490	Art Library (artlib)	Main Campus (wellic)	Wellesley College (wc)	arres	Art Reserve
5	Art	arttt	Art Library	c609759d-907a-4f01-80a9-32beb159409e	Art Library (artlib)	Main Campus (wellic)	Wellesley College (wc)	arttt	Art Library
6	Art	artresshelf	Art Open Reserves	36d0ff5e-b47d-468d-b4ba-69be0e29e76b	Art Library (artlib)	Main Campus (wellic)	Wellesley College (wc)	n/a	n/a
7	Art	artov	Art Oversize	56b87c2e-1e5e-4084-81de-2056e7ab0005	Art Library (artlib)	Main Campus (wellic)	Wellesley College (wc)	artov	Art Oversize
8	Art	artpr	Art Periodicals	5f82a1d7-8e6e-4626-b94b-aa9ac821037d	Art Library (artlib)	Main Campus (wellic)	Wellesley College (wc)	artpr	Art Period'l
9	Art	artrf	Art Reference	5edae0ac-310b-4bfb-b69f-636b54b2b3bb	Art Library (artlib)	Main Campus (wellic)	Wellesley College (wc)	artrf	Art Ref
10	Art	artca	Art Special Collections	e459f897-7177-42ee-885c-c1f741443c82	Art Library (artlib)	Main Campus (wellic)	Wellesley College (wc)	artca	Art Case
11	Art	artco	Art Special Collections Oversize	0c24ca40-3f7a-40d2-acb3-6720617b52c8	Art Library (artlib)	Main Campus (wellic)	Wellesley College (wc)	artco	Art Case Oversize
12	Art	artvd	Art Video	4f952291-72c9-4b50-97b3-65fe041d4d24	Art Library (artlib)	Main Campus (wellic)	Wellesley College (wc)	artvd	Art Video
13	Clapp	xastro	Astronomy Collection - Observatory	6aa70234-d127-4b95-836c-bfd5695ff8ee	Astronomy Collection (astrocoll)	Main Campus (wellic)	Wellesley College (wc)	astro	Astronomy
14	Clapp	clatl	Clapp Atlases - 2nd Floor Reference	e7b7f4bd-1f55-4133-a45b-5b9172d8c4da	Clapp Library (clapplib)	Main Campus (wellic)	Wellesley College (wc)	clatl	Clapp Atlas
15	Clapp	clact	Clapp Cataloging Department	5c29f00b-b30c-4f37-b9e7-f29acee66984	Clapp Library (clapplib)	Main Campus (wellic)	Wellesley College (wc)	clact	Clapp Acq/Cat Dept
16	Clapp	bindery	Clapp Conservation - bindery	b39db87b-4932-4c19-9254-c07b540b672b	Clapp Library (clapplib)	Main Campus (wellic)	Wellesley College (wc)	n/a was a status	
17	Clapp	inrepair	Clapp Conservation - in repair	17530099-1002-4eba-a1e3-1baaf19b7387	Clapp Library (clapplib)	Main Campus (wellic)	Wellesley College (wc)	n/a was a status	



FOLIO Call Number Types

	A	B	C
1	FOLIO Call Number Type	FOLIO UUID	Locations in Use
2	Dewey Decimal classification	03dd64d0-5626-4ecd-8ece-4531e0069f35	SpCol, Remote Storage
3	LC Modified	512173a7-bd09-490e-b773-17d83f2b63fe	
4	Library of Congress classification	95467209-6d7b-468b-94df-0f5d7ad2747d	SpCol, Clapp, Music, Art, Remote Storage
5	MOYS	828ae637-dfa3-4265-a1af-5279c436edff	
6	National Library of Medicine classification	054d460d-d6b9-4469-9e37-7a78a2266655	
7	Other scheme	6caca63e-5651-4db6-9247-3205156e9699	SpCol
8	Shelved separately	cd70562c-dd0b-42f6-aa80-ce803d24d4a1	
9	Shelving control number	28927d76-e097-4f63-8510-e56f2b7a3ad0	
10	Source specified in subfield \$2	827a2b64-cbf5-4296-8545-130876e4dfc0	
11	Superintendent of Documents classification	fc388041-6cd0-4806-8a74-ebe3b9ab4c6e	
12	Title	5ba6b62e-6858-490a-8102-5b1369873835	Rec Reading
13	UDC	d644be8f-deb5-4c4d-8c9e-2291b7c0f46f	
14	Unmapped	f763beca-9510-445b-b571-8cbfc2b42571	



Free Tools & Resources

- Google Colab (Colaboratory) for running Python scripts
- Python for Everyone (free ebook, available in multiple languages)
- Postman
- MarcEdit
- Visual Studio Code
- FOLIO's Data Import app
- OpenRefine



Postman & APIs

- [FOLIO Search Documentation](#) and [FOLIO API Documentation](#)
- Helpful to learn basics of APIs in conjunction with Network console in browser developer tools
- Primarily used GET calls to:
 - test endpoints
 - see how metadata was structured in JSON
 - test queries
- These tests help inform how we structured queries in Google Colab
- MARC query API

TIP: Each Endpoint

- will return a different maximum number of records
- will return different data fields



Google Colab & Python Scripts

- Used the [folioclient](#) python wrapper
- Used to identify records that needed changes and then implemented changes in batch
- Queries are written in CQL, the language FOLIO uses for search

TIP: create a test record that has every FOLIO field filled and an item for every status

- Sample query:
 - Using the holdings-storage endpoint to get thousands of records
 - Asks for all the records in Art Periodicals (location) with a holdings type of “Unmapped”

```
1 endpoint = "/holdings-storage/holdings"
2 record_type = "holdingsRecords"
3 page_size = "1000"
4 query = '?query=cql.AllRecords=1 AND permanentLocationId==5f82a1d7-8e6e-4626-b94b-aa9ac821037d AND holdingsTypeId==4effa3a8-c991-4882-b1ee-2ea7c7487820'
```



Google Colab & Python Scripts

- Update the Holdings Type example:
 - Change the Holdings type to “Serial”

UPDATE holdings type on holding record

```
1 updated_holdings = []
2
3 for holdings in data:
4     holdings["holdingsTypeId"] = "e6da6c98-6dd0-41bc-8b4b-cfd4bbd9c3ae"
5     updated_holdings.append(holdings)
6 # Electronic - 996f93e2-5b5e-4cf2-9168-33ced1f95eed
7 # Physical - 0c422f92-0f4d-4d32-8cbe-390ebc33a3e5
8 # Serial - e6da6c98-6dd0-41bc-8b4b-cfd4bbd9c3ae
9
10 #print(*updated_holdings[:100], sep="\n") the brackets say how many you will update out of the total; remove to do all; 5000 too many, 3000ok
11 print(*updated_holdings[:3000], sep="\n")
```



MarcEdit

- Used to edit MARC Bibliographic records for FOLIO's Data Import app
- Tasks for clean-up:
 - OCLC numbers
 - Add, change, or delete fields
 - Delete duplicate data
- Include or exclude records from a specific file
- Convert character encoding from MARC8 to UTF-8
 - FOLIO uses UTF-8 character encoding
- Create a Tab Delimited file (.csv) from a file of MARC records





FOLIO's Data Import App

- Limitations of Data Import
 - Multiple Holdings and Item records will fail = currently not supported
 - Relies on a local system identifier (HRID or UUID) for the best and fastest overlays and updates
- Created local statistical codes for retention agreements, items requiring electronic bookplates, items purchased with special donor funds
- Example: Eastern Academic Scholars' Trust (EAST) retentions
 - Identifying these records in FOLIO using Inventory's Query Search
 - (notes.note all "Retained for Eastern Academic Scholars' Trust (EAST)")
 - Add EAST statistical code to Instance, Holding, Item records
 - "Retention Agreement: EAST - EAST - DO NOT WITHDRAW"
 - Populate the retention policy field in holdings records
 - "EAST retention : Retain until 20310630"



Instance Update Job Profile

^ Overview

 Match profile: "001 to Instance HRID" ·  Instance · Instance → Instance HRID

^ For matches

 Action profile: "EAST - Instance Record Overlay" ·   Update instance

^ For non-matches

This section contains no items



Holdings Update Job Profile

 Match profile: "Holdings - 903ff\$h to Holdings HRID" ·  Holdings · Holdings → Holdings HRID

^ For matches

 Action profile: "EAST - Holdings Record Update" ·   Update holdings

^ For non-matches

This section contains no items



Item Update Job Profile

 Match profile: "Item - 949\$ to barcode" ·  Item · Item → Barcode

^ For matches

 Action profile: "EAST - Item Record Update" ·  Update item

^ For non-matches

This section contains no items



Managing clean-up a year later...

- Some of the clean-up tasks we completed between July-December 2022 can now be done with the Bulk Edit app
 - New functionality comes with each flower release
- Bibliographic Record Authority Control
- We have tasks that were lower priority that we may complete this fall and winter
 - Adding the call number types to holdings records
 - Changes needed for our library renovation (ex. locations, item status updates, notes)
- We have shifted our focus to learning SQL with LDP for generating reports.
 - Closest we have found to Millennium's "Create List" feature



I need help! Where do I go?

- Implementation Consultants for initial months post-migration
- FOLIO Slack Community
 - Data Import Lab Sessions
 - Learning APIs
 - Metadata Management SIG
- FOLIO Wiki
- FOLIO Search documentation
- W3 Schools



Lessons Learned

- Don't struggle on your own!
 - There are no silly questions. Just ask!
- Get involved in the SIGs prior to your migration (when you have time).
 - Ask your colleagues to attend the SIGs relevant to their work
- Do as much as possible in your legacy system as possible
- Get as much RAM for your computer as IT will let you. Trust us, it's worth it.
- LDP/Metadb is worth it for getting reports.
- Learning together is helpful.
- Visit other FOLIO libraries' githubs

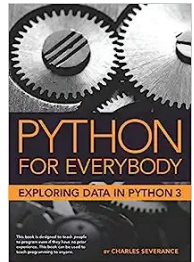


Tools & Resources

- [Google Colab](#)
- [Python for Everybody](#)
- [Postman](#)
- [MarcEdit](#)
- [Open Refine](#)
- [Visual Studio Code](#)
- [FOLIO Search documentation](#)
- [FOLIO API Documentation](#)
- [Working with FOLIO APIs](#)
- [folioclient \(Python Library\)](#)
- [LDlite](#)
- [LDP](#)
- [Dbeaver](#)



Visual Studio Code



Questions?

Lynne Fors

Cataloging & Metadata Librarian

lf103@wellesley.edu

Kara Hart

Systems Librarian

khart@wellesley.edu

For this presentation and other resources:

https://github.com/wellesleyfolio/WOLFcon_2023

