



Publishing without Perishing

Alma Publishing Profiles for Fun and Profit

ELUNA 2023

Martin Patrick: martin.patrick@library.gatech.edu

Stacie Trill: trail001@umn.edu



Publishing Profiles 101



The Basics

- Slightly text heavy slides today, more in the speaker notes and appendix
- Data export options in Alma
- Pub profiles background and use cases



The Basics: Alma Export Jobs

Export Authority records

Export Bibliographic records

Export Digital titles

Export Electronic portfolios

Export Physical items

But how to combine them all?



The Basics: Alma Analytics Exports



PDF



Excel



Excel



CSV



Powerpoint



Tab Delimited



Web Archive



XML

You could take tabular data and use a tool like MARCEdit to turn it into MARC records, or you could use pymarc to explore it



Pros and Cons of Jobs and Analytics Export

Pros

- Data comes out exactly as it
- Basic setup to export from Alma
- If you want just MARC data, tabular data, or MARCXML

Cons

- Cumbersome to combine bibs and holdings and items, etc.
- Data comes out exactly as it is
- Lack of output targets
- Analytics doesn't have every field indexed and requires several weird tricks to combine things

A decorative horizontal bar consisting of a teal segment on the left and an orange segment on the right.

The Basics: Alma Publishing

- Not just for sending data to Primo or BrowZine
- Flexible output options: FTP servers or OAI
- Ability to normalize data on the fly
- Bibliographic data is customizably enrichable with holdings, items, and portfolios
 - Choice of publishing one bibliographic for each item, or one bibliographic record with all attached items

A decorative horizontal bar consisting of a teal segment on the left and an orange segment on the right.

Publishing Profiles: Best Practices (ioho)

- Publishing On option
- Publishing Order of Operations
- Enrichment Options
- Filtering
- Normalizing

The Basics: Publishing Settings

Content

Set name * ☰ 🕒

Additional set name ☰

Filter out the data using ▼ ℹ️

Publish on:

- Bibliographic level
- Campus level
- Holdings/Portfolio/Representation level
- Item/Portfolio level

Output format ▼

Publishing Protocol

- MARC21 Bibliographic
- Dublin Core (Simple)
- FTP Dublin Core (Qualified)
- ETD-MS
- OA BIBFRAME Bibliographic
- RDA/RDF
- Z39.50



The Basics: Alma Publishing Enrichment

Bibliographic Normalization

Correct the data using normalization processes

Linked Data enrichment

Bibliographic Enrichment

Add Management Information

Related Records Enrichment

Add related records information

Authority Enrichment

Add Authority Information

Physical Holdings Enrichment

Add Holdings information

Physical Items Enrichment

Add Items Information

Committed to retain subfield

Retention note subfield

Electronic Inventory Enrichment

Add Electronic Portfolio Information

Digital Inventory Enrichment

Add Digital Representation Information

Add File Information

Add Remote Representation Information

Collection Enrichment

Add Collection Information

Analyzing Bibliographic and Item Data: The Mysterious Case of the Retention Notes



Background

- Retention notes for HathiTrust and Scholar's Trust had:
 - Been added to 583 in the Bibliographic Record in all cases
 - Been added to the Holdings record in some cases
 - Had the Retention Commitment flag turned on in zero cases
- The list of which actual items by barcode were committed was not made available, however I was assured the bib records were accurately marked. 😊



Questions to address

- How many of the bib records had more than one holdings record?
- How many holdings records had more than one item?
- Which holdings records *didn't* have a 583 note?
- From a set of bib records, how can I best get a set of barcodes so I can update the Retention Flag and notes in the correct item record itself?



Setting up Publishing

Physical Items Enrichment

Add Items Information

Repeatable field *

Item PID subfield

Copy ID subfield

Item Policy subField

Barcode subfield

Material type subfield

Item status subfield



Processing the Output



OCLC Publishing

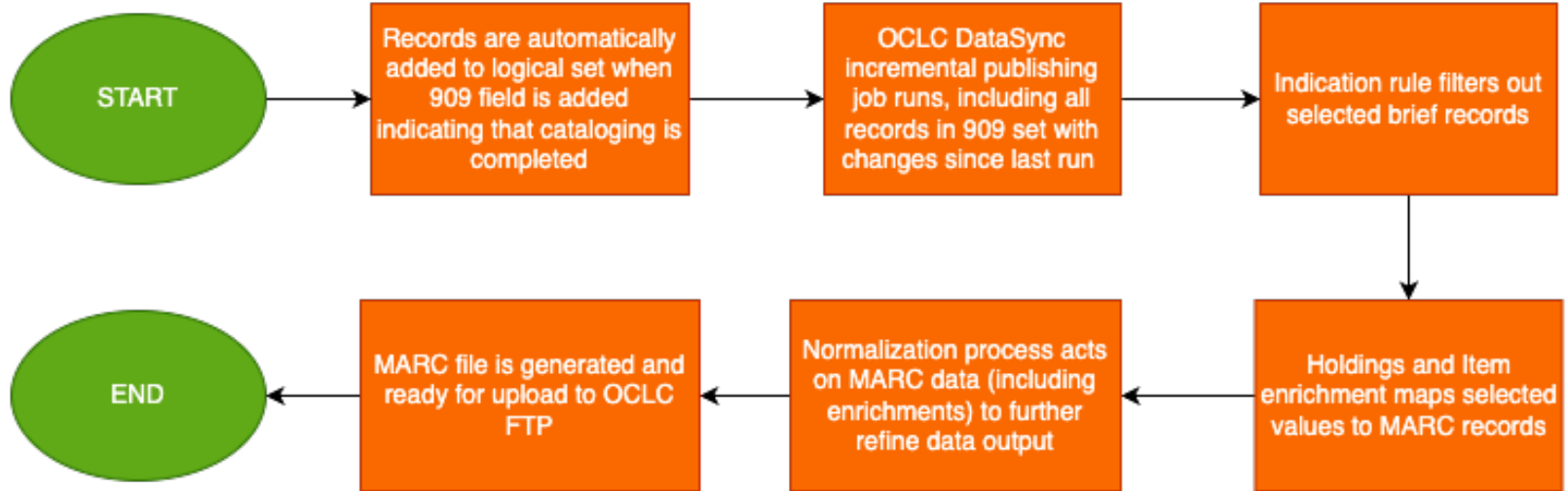


Why would you want to?

- ~~You need to set holdings for multiple symbols at a location level~~
- You wish to customize data output in ways not possible through out-of-the-box OCLC publishing.
- You want to control which records are sent to OCLC by means other than the Export to WorldCat flag (using logical set creation and/or filtering in publishing).
- You want to publish to more than one OCLC DataSync collection.
- You want more control over OCLC holdings deletion than the out-of-the-box job offers.



Process Flow





Sets and Filtering

- Publishing to OCLC is managed on the basis of a local field defined to record **Cataloging Status** (MARC 909).
- No record is published to OCLC unless its MARC 909 indicates that cataloging has been completed.
- A logical set based on 909 data is the basis of incremental publishing.

```
909  __ |a cat |b MNU |c 2023-04-04
```

Set name OCLC Datasync July 2022 and onward with OCN **Set type** logical
where (Other System Number contains keywords "OCoLC" AND
c.search.index_names.local_field_909 contains keywords "2022*") OR (Other System Number
contains keywords "OCoLC" AND c.search.index_names.local_field_909 contains keywords
"2023*")



Sets and Filtering

- Some brief records for special projects that wouldn't otherwise be excluded from the logical set are excluded on the basis of an indication rule in the publishing profile.

```
rule "909 brief projects"  
when  
  (exists "909.d.iso_std" OR exists "909.d.nr1_g1")  
then  
  set indication."true"  
end
```



Customizing Data Output

- Normalization routine for OCLC DataSync includes rules to:
 - Delete local fields.
 - Delete OCLC symbol mapping fields for records carrying a locally-designated “no export” field.
 - Map locations in a statewide storage facility to the appropriate OCLC symbols.
 - Remediate some specific problems with Alma MARC output that have caused issues with OCLC in the past (010, 880).
- Normalization happens **after** holdings and item enrichment, so normalization can act on the basis of data added through those enrichment processes.



Publishing to more than one DataSync collection

- U of MN uses separate OCLC DataSync collections for records with OCLC numbers and records without OCLC numbers in order to use different matching algorithms.
 - Records with OCLC numbers match on that basis only
 - Records without OCLC numbers use OCLC's full match algorithm
- There is one logical set and one general publishing profile for each DataSync collection.
- Both publishing profiles use the same filtering and data customizations.

Profile Details

Profile Name * **OCLC Datasync mid-2022 on: physical records without OCNs**

Profile Description **Publishing profile for OCLC Datasync. Incremental publishing for records without OCNs and 909 dates 6/20/22 and onward. Files published by this profile are for Datasync collection ID 1025409 which uses OCLC's full match algorithm.**



Holdings Deletes

- Holdings deletes are handled separately, through Alma Analytics reports.
- This process makes it easier to ensure that holdings are only deleted for the symbols corresponding to libraries who have withdrawn their last copy of a title.
- Delete reports are fetched on a schedule by a script through the Analytics API and uploaded to OCLC's FTP.



Producing metadata for external repositories

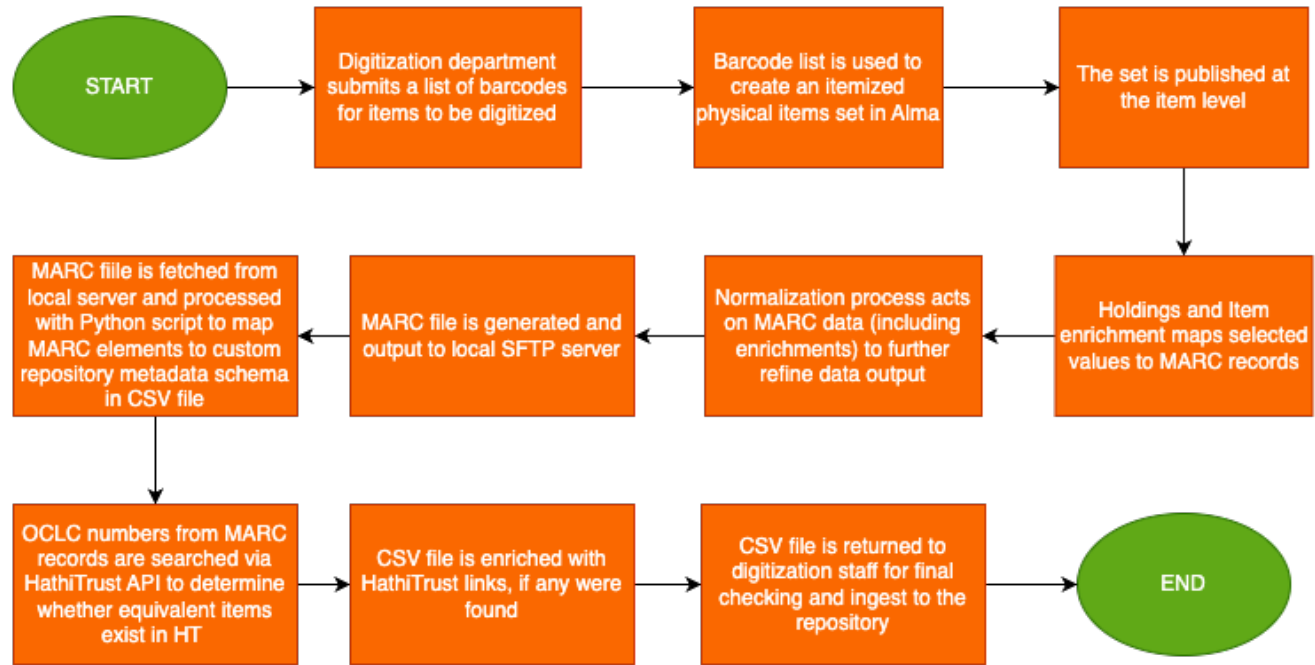


Background

- Locally held items are digitized and made accessible through an instance of ContentDM using a homegrown metadata schema
- **Goal:** extract and transform MARC records for physical items to metadata suitable for ingest to ContentDM
- Multi-step workflow using Alma Publishing + Python allows transformations and enrichments at various points throughout the process.



Process Flow





Enrichment: Holdings and Item Data

- The publishing profile produces item-level output (1 record per item).
- Holdings and item data needed in the repository is added to records through Physical Holdings Enrichment in the publishing profile.

Physical Items Enrichment

Add Items Information

Repeatable field	955	Barcode subfield	a
Item PID subfield		Material type subfield	
Copy ID subfield		Item status subfield	
Item Policy subfield		Is Magnetic subfield	
Provenance subfield		Enumeration B subfield	
Enumeration A subfield		Chronology J subfield	
Chronology I subfield		Process type subfield	
Description subfield	v	Permanent location subfield	c
Permanent library subfield	b	Current location subfield	
Current library subfield		Call number subfield	d
Call Number type subfield			



Transformations, part 1: Alma Normalization Rules

- An Alma normalization routine was created which is used only in the publishing profile for this workflow.
- The NR removes fields not needed in the repository metadata and remaps some fields to alternative tags to make further transformations easier. For example, subjects are handled with less granularity in the repository than they are in MARC, so:

```
rule "UMedia subjects mapping"  
when  
  (TRUE)  
then  
  copyField "600" to "699.{0, }"  
  copyField "610" to "699.{1, }"  
  copyField "611" to "699.{1, }"  
  copyField "650" to "699.{5, }" if (exists "650.{ ,0}")  
  copyField "651" to "698"  
end
```



Transformation example: subject fields

Alma original record:

```
600          10 |a La Salle d'Offémont,  
Adrien Nicolas Piédefer, |c marquis de,  
|d 1735-1818
```

```
650          _0 |a Slavery |z Haiti |v  
Early works to 1800.
```

```
650          _0 |a Black people |z Haiti  
|v Early works to 1800.
```

Published transformed record:

```
699  0\ $aLa Salle d'Offémont, Adrien  
Nicolas Piédefer,$cmarquis de,$d1735-  
1818
```

```
699  5\ $aSlavery$zHaiti$vEarly works to  
1800.
```

```
699  5\ $aBlack people$zHaiti$vEarly  
works to 1800.
```



Transformations, part 1: Alma Normalization Rules

- Another remapping example: non-Latin script titles, authors, and contributors are mapped from 880 fields to their corresponding MARC tags:

```
rule "UMedia 880 mapping"  
when  
  (TRUE)  
then  
  copyField "880" to "246.{ , }" if (exists "880.6.245*")  
  copyField "880" to "246.{ , }" if (exists "880.6.246*")  
  copyField "880" to "190.{1, }" if (exists "880.6.100*")  
  copyField "880" to "191.{2, }" if (exists "880.6.110*")  
  copyField "880" to "700.{1, }" if (exists "880.6.700*")  
  copyField "880" to "710.{2, }" if (exists "880.6.710*")  
end
```



Transformation example: non-Latin script fields

Alma original record:

880 10\$6245-02\$a本草圖譜 :\$b[93卷及索引] /\$c灌園岩崎常正著 ; 編纂者 飯田藏太郎.

880 1\1\$6100-01\$a岩崎灌園,\$d1786-1842.

880 1\1\$6700-06\$a飯田藏太郎.

Published transformed record:

246 \1\$6245-02\$a本草圖譜 :\$b[93卷及索引] /\$c灌園岩崎常正著 ; 編纂者 飯田藏太郎.

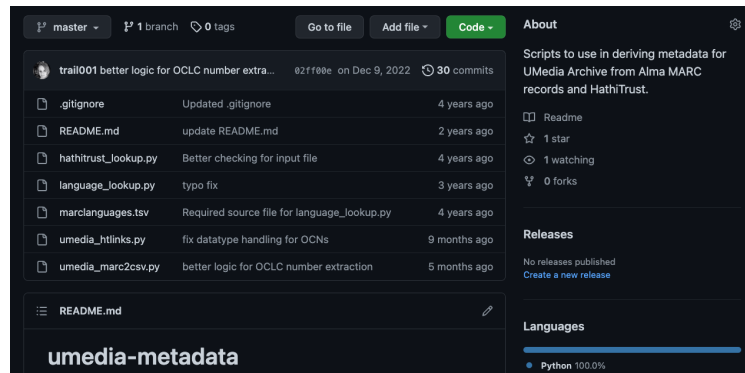
190 1\1\$6100-01\$a岩崎灌園,\$d1786-1842.

700 1\1\$6700-06\$a飯田藏太郎.



Transformations, part 2: MARC to CSV with Python

- The customized MARC output is processed by a Python script that outputs a CSV file suitable for ingest to the repository.
- A second script further enriches the CSV file by adding HathiTrust links when there is an OCLC number match.





Transformations, part 2: MARC to CSV with Python

```
#Subjects mapping
if field_699:
    subj_list = []
    for f699 in field_699:
        subjs = f699.get_subfields('a', 'x', 'y')
        for subj in subjs:
            subj_list.append(subj.rstrip('.'))
    subjects = '||'.join(subj_list)
    rec_dict['Subject'] = subjects
```

Subject
La Salle d'Offémont, Adrien Nicolas Piédefer Slavery Black people
Slavery Early works to 1800
Slavery History 18th century Slavery History 18th century

Fin

Martin Patrick: martin.patrick@library.gatech.edu

Stacie Trill: trail001@umn.edu

Appendix



Analyzing OAI Feed with MARCEdit



Why would you even want to?

Some use cases:

- Ensuring your RapidILL/BrowZine feeds are accurate
 - Norm rules are working, things you can't lend are being properly excluded, etc.
- If you don't have an FTP server or don't want to use Ex Libris Secure FTP
 - (or can't for because it's a whole thing with IT)



Ok, so what do I need?

What you need:

- Access to the Integration Profiles in Alma Configuration, or a willing Systems Librarian
- [OAI_DEFINITION enabled](#)
- MARCEdit installed (Mac or Windows)
- A publishing profile that has published
- The profile's set name
- Your current IP address
- Your OAI Repository Base URL

< Integration Profile

General Information

Actions

Contact Info

OAI REPOSITORY DEFINITIONS

Active * Active Inactive

Repository Name *

Repository Base URL

Protocol Version

Admin Email *

Earliest Datestamp *

Date Format is dd/MM/yyyy

Deleted Record

Granularity

[Metadata Prefixes](#)

[Allowed IPs](#)

	Enabled	IP description	IP Version	IP Match Criteria	Updated By	Last Updated	
1	<input checked="" type="checkbox"/>	Third Iron IP	IPv4	[REDACTED]	903284307	11/13/2019	...
2	<input checked="" type="checkbox"/>	Martin P testing	IPv4	[REDACTED]	903862682	09/07/2022	...
3	<input checked="" type="checkbox"/>	RapidILL 1	IPv4	[REDACTED]	903862682	09/07/2022	...
4	<input checked="" type="checkbox"/>	RapidILL 2	IPv4	[REDACTED]	903862682	09/07/2022	...
5	<input checked="" type="checkbox"/>	Martin P Testing 2	IPv4	[REDACTED]	903862682	04/20/2023	...

MarcEdit 3.5.81 by Terry Reese

Tools ▾ Plugins ▾

- Log File Management >
- Delimited Text Translator
- Export... >
- Character Conversion Tools >
- MARC SQL Explorer
- Harvest OAI Records**
- Generate Reports >
- OpenRefine >
- MARC Format Conversion
- Verify URLs
- Virtual Keyboard ⌘ ⌘ V
- Z39.50 Client
- MARC Processing Tools >
- OCLC Operations >
- Linked Data Tool
- Validate Headings >

stering Tools

imited Text Translator

Helper

RCValidator

vest OAI Records

.50/SRU Client



Metadata Harvester

Home Advanced Settings

Server Address

Set Name

Metadata Type

Crosswalk Path:

Start:

End:

Use System Proxy

=LDR 00916pam a22002534a 4500
=005 20000830133008.0
=008 000303t20002000nyua\\\\b\\\\001\\0\\eng\\
=001 995010253402947
=010 \\\$a 00028512
=020 \\\$a3540667997\$qsoftcover\$qalkaline paper
=035 \\\$a(OCOLC)ocm43648596
=040 \\\$aGAT\$beng\$cGAT
=042 \\\$apcc
=049 \\\$aGATT
=050 00\$aQ337\$b.C65 2000
=245 00\$aComputational conflicts :\$bconflict modeling for distributed intelligent systems, with contributions by
numerous experts /\$cHeinz Jürgen Müller, Rose Dieng (eds.).
=264 \1\$aNew York :\$bSpringer,\$c[2000]
=264 \4\$c©2000
=700 1\$aMüller, Heinz Jürgen.\$0<http://id.loc.gov/authorities/names/n00003024>
=700 1\$aDieng, Rose.\$0<http://id.loc.gov/authorities/names/n00003025>
=945 \\\$aMARCIVE20210914
=950 \\\$agatech
=994 \\\$aE0\$bGAT
=954 \\\$rGENERAL\$tQ337 .C65 2000\$b50671011859706\$qGTLSC



RapidILL Norm Rule Examples

Martin's RapidILL Norm Rules: <https://github.com/martinpatrick/rapid-norm-rules>