

Discussion:
**Library-style Authority Control
in ArchivesSpace**

Kevin Schlottmann
Head of Archives Processing
Rare Book and Manuscript Library
Columbia University

ArchivesSpace Member Forum
August 14, 2018

Columbia plans to have a library vendor perform regular authority work on archival records.

We are exploring workflows and scripts that we are going to share out the community, but would also love to discuss how others are approaching authority control in a mixed library/archival description environment.

Discussion

Many of us are doing this work

Still lots of work to do

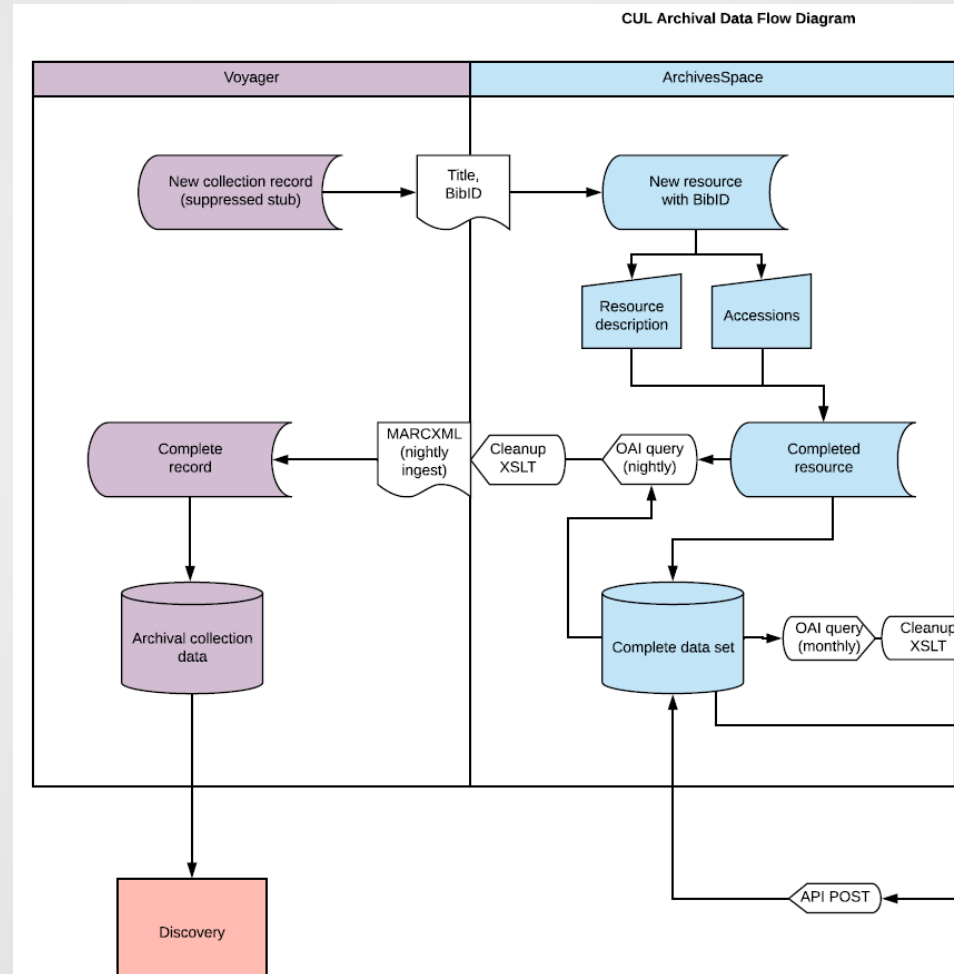
Please ask questions and provide comments

Why?

Context: Migration to AS

- Exporting collection-level AS records into Voyager (ILS) and then into discovery portal CLIO
- Upshot - records need to undergo authority control to be searchable and facetable

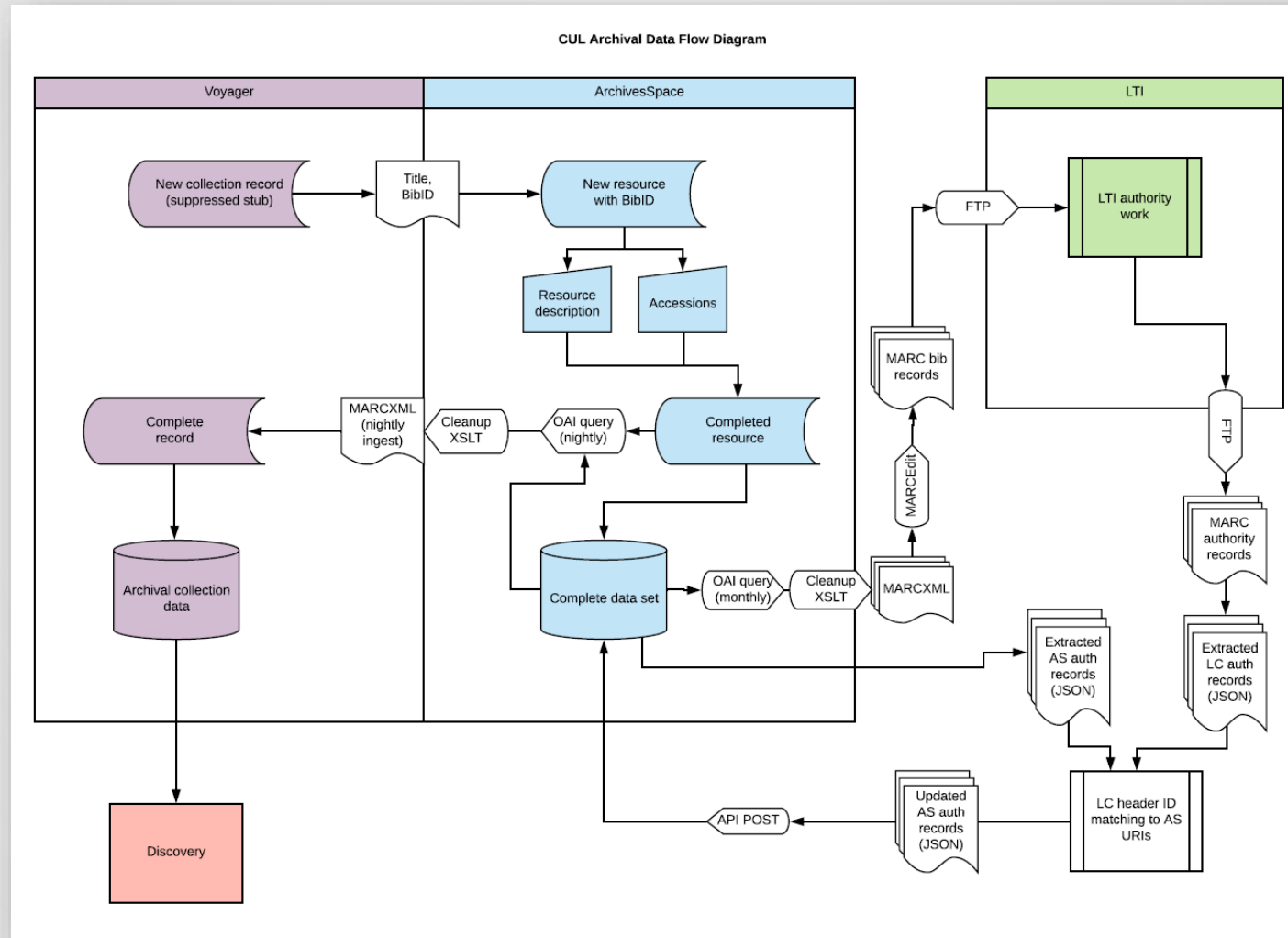
Nightly Data Flow



Authority Work

- Considered two-way ILS-AS data flow
 - Too expensive to develop and maintain
 - Invests in ILS which will not be around forever
- AS is the record of record, and data flows out

How?



How?

These scripts are inspired by and indebted by many existing ASpace code repos, but in particular to Lora Woodford's API workshop (<https://github.com/lorawoodford/api-training>).

<https://github.com/cul/rbml-archivesspace/tree/master/authorityControl>

How?

- Key: LC identifiers to agent records claiming to be LC
 - Why “claiming”? The import included decades of legacy description

How?

- Add LC identifiers to agent records claiming to be LC
 - Export agent records in JSON
 - Call the LC API to match on strings
 - Human review
 - Write the updated agent records back in

<https://github.com/cul/rbml-archivesspace/tree/master/authorityControl>

First Pass

Filtered By Clear All

- Record Type: Person ✕

Filter by text

Source

- NACO Authority File 6673
- Unspecified ingested source 4462
- Library of Congress Subject Headings 1946
- Local sources 45

Rules

- Local rules 39
- Anglo-American Cataloging Rules 1

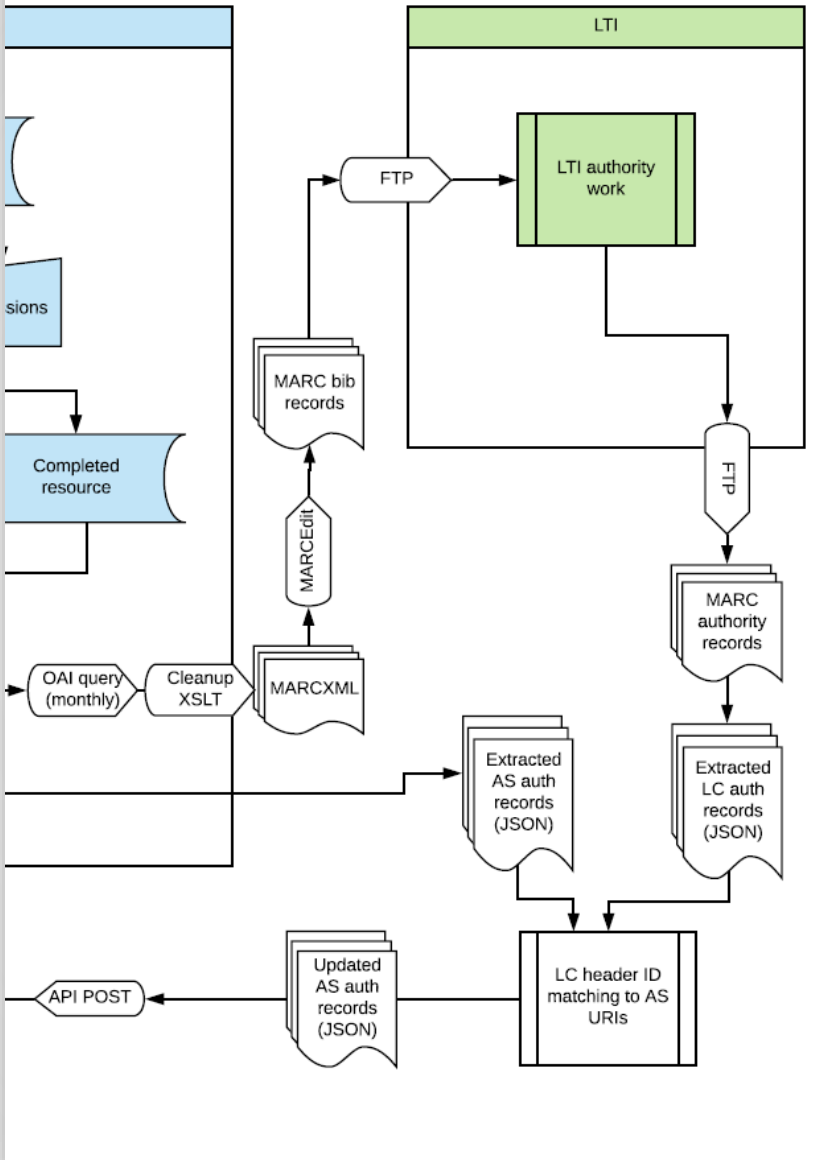
Edit Required Fields ▾ Edit Default Values ▾ Download CSV Create Agent ▾ Delete

Agents

Showing 1 - 30 of 13126 Results, Sort by: Agent Name Ascending ▾ and: Select ▾

	Record Type	Agent Name ▾	Authority ID	Source	Rules	
<input type="checkbox"/>	Person	Å Beckett, Arthur William, 1844-1909	http://id.loc.gov/authorities/names/n82155066	NACO Authority File	Created by admin 2018-06-03 15:32:51 UTC Last Modified by kws2126 2018-08-09 19:34:40 UTC	View Edit
<input type="checkbox"/>	Person	Aalto, Alvar, 1898-1976	http://id.loc.gov/authorities/names/n79018877	NACO Authority File	Created by admin 2018-06-01 17:38:20 UTC Last Modified by kws2126 2018-08-09 19:30:22 UTC	View Edit
<input type="checkbox"/>	Person	Aalto, Elissa		Unspecified ingested source	Created by admin 2018-06-01 17:55:43 UTC Last Modified by admin 2018-06-01 17:55:43 UTC	View Edit

Flow Diagram



Next...

Next

- 1) API call to ASpace to get fresh extract of json files on all subjects and agents
(getAgents.py)
- 2) Send bib records to vendor via OAI feed
(cleaned up via cleanOAI.xsl)

<https://github.com/cul/rbml-archivesspace/tree/master/authorityControl>

Next

3) From returned MARC *authority* records, extract csv with authno, primary name, sort name, anything else we want from the authority records (MARCedit, or pymarc, or some string manipulation; not trivial but doable and scriptable)

<https://github.com/cul/rbml-archivesspace/tree/master/authorityControl>

Next

4) Script to match csv to json record on authority number, and write in primary name, sort name, etc. Only post modified json files back in.

(modification of postAgents.py)

<https://github.com/cul/rbml-archivesspace/tree/master/authorityControl>

New Records

For newly established agents, archivists manually insert the LC link.

What's Left

1) Actually doing the vendor round-trip

2) Legacy data:

The screenshot shows a web interface with a filter bar at the top. The filter bar includes a 'Filtered By' section with a 'Clear All' link and a list of filters: 'Record Type: Person' with a close icon. Below the filter bar is a search box labeled 'Filter by text' with a magnifying glass icon. The main content area is divided into two sections: 'Source' and 'Rules'. The 'Source' section lists four categories with their respective counts: 'NACO Authority File' (6673), 'Unspecified ingested source' (4462), 'Library of Congress Subject Headings' (1946), and 'Local sources' (45). The 'Rules' section lists two categories: 'Local rules' (39) and 'Anglo-American Cataloging Rules' (1). Two red arrows point from the right side of the image to the 'Unspecified ingested source' and 'Library of Congress Subject Headings' entries.

Filtered By Clear All

- Record Type: Person ✕

Filter by text

Source

- NACO Authority File 6673
- Unspecified ingested source 4462
- Library of Congress Subject Headings 1946
- Local sources 45

Rules

- Local rules 39
- Anglo-American Cataloging Rules 1

What's Left

Discussion and development!

<https://github.com/cul/rbml-archivesspace/tree/master/authorityControl>

<https://github.com/archivesspace/awesome-archivesspace>

@archivistkevin

kws2126@columbia.edu