Collaborative automation across systems:

standardizing records and workflows with the ArchivesSpace API

Jessica Geiser, Collections Management Librarian

Noah Geraci, Digital Assets Metadata Librarian

UCR | LIBRARY

How do you use a clean-up project to make lasting process improvements?

Project: Initial Goals

Our initial goals for the project

- Eliminate "hidden collections"

- Update and standardize language across collections

- Create consistent records across all access points

5

Title   Mexico photograph collection.

	Connect to:
	Finding aid:

LOC	CALL #	STATUS
SpCol Archives	**MS 022**	LIB USE ONLY ⚲ Map it

Send via SMS/Email   Help

⌂ Permanent Link: http://scotty.ucr.edu/record=b1014212~S5

Call #	MS 022
Description	1.17 linear feet (2 boxes, 2 photograph albums)
Content Type	still image
Media Type	umediated
Carrier Type	sheet
Notes	Processing Information: This collection is not fully processed.
Access	Please contact Special Collections & Archives in advance to request access.
Notes	In English and Spanish.
Summary	This collection consists of photographs and stereoscopic photographs showcasing the landscapes, people, and culture of Mexico. Includes images of Texcotzingo, Miraflores, Coyoacán, and Tlalmanalco.
Preferred Citation	Mexico photograph collection, MS 022. University of California, Riverside Libraries, Special Collections & Archives, University of California, Riverside.
Publication/Use Rights	Copyright has not been assigned to the University of California, Riverside Libraries, Special Collections & Archives. All requests for permission to publish or quote from manuscripts must be submitted in writing to the Head of Special Collections & Archives. Permission for publication is given on behalf of the Regents of the University of California as the owner of the physical items and is not intended to include or imply permission of the copyright holder, which must also be obtained by the researcher.
Subject	**Mexico -- Photographs.**
Genre/Form	**Mexico.** **Photographs. aat** **Stereoscopic photographs. aat** **Photographs. fast (OCoLC)fst01423689**
Other #	(OCoLC)17852654
OCLC/BIB	17852654

Title   Harry W. Lawton collection on Willie Boy.

Produced   bulk 1889-1996.

LOC	CALL #	STATUS	Request	
SpCol Archives	**MS 152**	LIB USE ONLY	Special Collections Request	⚲ Map it

Send via SMS/Email   Help

⌂ Permanent Link: http://scotty.ucr.edu/record=b4738567~S5

Call #	MS 152
Description	8 linear feet
Content Type	text
Media Type	unmediated
Carrier Type	sheet
Summary	Consists of Harry W. Lawton's research pertaining to Willie Boy, the Paiute Indian outlaw of Southern California. Correspondence, documents, and press clippings are included. It also contains material related to Lawton's book Willie Boy: A Desert Manhunt and the lawsuit for libel filed by Lawton against the authors of The Hunt for Willie Boy: Indian Hating and Popular Culture.
Other #	(OCoLC)900884853
OCLC/BIB	900884853

6

Collaboration

Why automate?

- Bulk updates

- Time and resources involved

- Providing easy solutions for future updates

- Stylesheet fixes

Current Steps for Updating Collection Level ASpace Record

In ASpace

- Add Resource Type from drop-down
- Take period out from date statement
- Make sure the date type is correct, add years under type
- Check extent
- Remove type of boxes from container summary, put in parenthesis
- Take out "Guide to the" from Finding Aid title
- Remove Finding Aid author
- Check agents against LCNAF – edit if needed
- Check subjects, adding or editing as needed
  - Put in alphabetical order by subject, then genre/form (as it would appear on OAC)
- Basic note checks – check Abstract, Languages, Scope and Contents to make sure they're accurate enough – add them if needed.
- Change Access note to:
  - This collection is unprocessed. Please contact Special Collections & University Archives regarding the availability of materials for research use.
- Check Conditions governing use note:
  - Replace "Head of Special Collections" with "Director of Distinctive Collections" where it appears
- Check Preferred Citation:
  - Should be [identification of item], [date if possible]. *Collection Title* (*Collection Identifier*). Special Collections & University Archives, University of California, Riverside.
- Delete Processing History note
- Scope and Contents Note Label should be "Collection Scope and Contents"

For OAC

- Download EAD file from ASpace record
- Follow EAD modifications guidelines to edit xml
- Upload to OAC Test space
  - Check record online to ensure it's correct

EAD modification/validation using <oXygen/> XML Editor

The following steps should be taken on all ArchivesSpace generated EAD files before they are uploaded to the Online Archive of California (OAC).

1. Remove *<num>* tag in *<titleproper>*

2. Add the date after *<author>* in *<titlestmt>*

   (ex. <author>Finding aid prepared by Eric Milenkiewicz and Noah Geraci, 2017.</author>)

3. Replace the included *<publicationstmt>* with:

   <publicationstmt>

   <publisher>Special Collections & University Archives</publisher>

   <address>

   <addressline>The UCR Library</addressline>

   <addressline>P.O. Box 5900</addressline>

   <addressline>University of California</addressline>

   <addressline>Riverside, California 92517-5900</addressline>

   <addressline>Phone: 951-827-3233</addressline>

Technical approaches

Road map

```
┌────────────────────────────┐
│   Data remediation in      │
│      ArchivesSpace         │
└────────────────────────────┘
```

Bulk export EAD	Bulk export MARCXML
Transform EAD	Transform MARC
OAC ingest	OCLC + Millennium ingest
Repurpose code for ongoing use [command-line tool]	Repurpose code for ongoing use [TBD]

Requirement

Scope and Contents Note Label should be "Collection Scope and Contents".

Code

```python
if note['type'] == 'scopecontent':

    note['label'] = 'Collection Scope and Contents'
```

Requirement

Check agents against LCNAF – edit if needed.

Process

- Python script to export all agent names to CSV file.

- Use OpenRefine to reconcile names with LCNAF.

- Python script to read reconciled CSV and update ArchivesSpace agents (put URI in "Authority ID" field).

Tools

- **Python:** familiar, [an abundance of examples](#), libraries for xml (lxml) and MARC (pymarc)

- **XSLT:** handles XML structure well, utilize existing stylesheets for EAD, can be called from Python

- **MarcEdit:** used as adjunct tool for MARC conversion and validation in initial cleanup

Ongoing automation: needs

- Maintainable at the librarian level in a large academic library

- Usable across levels of technical interest and experience

Why a command-line tool?

- "Small pieces, loosely joined"

    - (simplicity, flexibility, maintainability)

- Manageable project for novice developer

- Worried about usability, but archives colleagues were game!

oac-process

oac-process takes one or more EAD files exported with ArchivesSpace defaults and tidies them up according to local guidelines. It validates the EAD to alert to any errors or issues. The resulting file is an EAD file ready for upload to OAC.

Its default behavior also moves the files to standard locations on the UCR Library shared drive.

Example:

```
oac-process UA.398_20180424_193147_UTC__ead.xml
```

Output:

```
EAD validated
ua398.xml processing completed
```

Lessons learned

- Significant improvements are attainable

- Best tool is one that works in your environment

- ArchivesSpace is the easy part

- Clarity of requirements and documentation are key

Thank you!

https://github.com/ngeraci/ucr_archivestools

jessica.geiser@ucr.edu

noah.geraci@ucr.edu