

University of Massachusetts Amherst
ScholarWorks@UMass Amherst

University Librarians Presentations Series

University Libraries

2019

Creating a Sustainable Workflow for Converting Online Finding Aids into MARC Records

Ann Kardos

Follow this and additional works at: https://scholarworks.umass.edu/librarian_presentations

 Part of the [Cataloging and Metadata Commons](#)

Kardos, Ann, "Creating a Sustainable Workflow for Converting Online Finding Aids into MARC Records" (2019). *Association of College and Research Libraries 2019 Conference*. 9.

Retrieved from https://scholarworks.umass.edu/librarian_presentations/9

This Article is brought to you for free and open access by the University Libraries at ScholarWorks@UMass Amherst. It has been accepted for inclusion in University Librarians Presentations Series by an authorized administrator of ScholarWorks@UMass Amherst. For more information, please contact scholarworks@library.umass.edu.

Creating a sustainable workflow for converting online finding aids into MARC records

UMass Amherst Libraries

Creating a sustainable workflow for converting online finding aids into MARC records

Ann Kardos
University of Massachusetts Amherst

What it was

All the University of Massachusetts Libraries all special collections and archival materials are described using the online finding aids by Special Collections and University Libraries (SCUL) staff. Everything has at least the usual description fields. It is difficult to process the collection (usually a student collection or similar). These staff records are manually transferred into the OAI-PMH Bag managed by SCUL. It is not and it is not intended as the primary searching tool.

The good & the bad

The Five Colleges (Amherst College, Hampshire College, UMass Lowell, UMass Dartmouth and the University of Massachusetts Amherst). The Five Colleges share a legacy of finding aids to patrons and many different databases. It is not clear if any of the Five Colleges can find any of the location of any records.

What it became

Over 1000 records have been identified in OAI-PMH and added to the shared catalog and discovery. This creates an additional avenue for student researchers to access UMass Amherst's unique collections. It also allows potential discovery by faculty and staff, members of the general public or interested parties. It may not mean that an item is open and available to use with access. Patrons must also have the means and interest in primary source materials or unique collections that they may not normally have had access to.

Workflow

Special Collections and University Libraries (SCUL) uses a Python script to scrape and ingest OAI-PMH and generate the records into OAI. The database is then used by the OAI. It is and begins the process.

Five College collaboration

- Five College Libraries will pass to OAI-PMH Bag Files and generate a standardized list of all the Five Colleges records special collections and archival materials available. The goal was for the Bag Files to come up with a standard that can be discovered as possible in the student environment, but with minimal effort because of a learning migration from OAI-PMH. Some of the Five Colleges cannot utilize OAI-PMH directly used OAI-PMH files and it was hoped that UMass Amherst could act as a proxy to OAI-PMH files as well.

The Bag Files are sought to define a standard OAI-PMH profile.

Ann Kardos

University of Massachusetts Amherst



PRESENTED AT:

WHAT IT WAS

At the University of Massachusetts Amherst, all special collections and archival materials get described when they come in the door by Special Collections and University Archives (SCUA) staff. Everything has at least that "stub" description while it awaits someone to process the collection (usually a student worker or volunteer). These stub records, and eventually finding aids, live in a WordPress Blog managed by SCUA. It's called UMarmot (<http://scua.library.umass.edu/umarmot/scua/what-is-umarmot/>) and it also functions as their in-house searching tool.

The screenshot shows the website for Special Collections and University Archives at UMass Amherst. The browser address bar indicates the URL is scua.library.umass.edu/umarmot/. The page has a red header with the UMass Amherst logo and a navigation menu with options: About SCUA, Collections, Outreach, and Research. A search bar is prominently displayed on the right side, with a dropdown menu showing categories: Manuscripts, Digital, and Books. Below the search bar, there are sections for 'Browse collections' (with a letter index A-Z and UMass), 'Collecting areas' (with a 'Select Category' dropdown), and 'UMass Archives' (with a 'Chancellors (2)' dropdown). The main content area features a large historical photograph of a group of people in early 20th-century attire. In the foreground of the photo, a barrel is visible with the text 'THE STRA...'. The footer of the page contains the text '© 2019, UMass Amherst | Site policies | Site contact' and a search bar with the text 'Find collections'.

SCUA collaborates with the Metadata Unit for anything that requires MARC metadata, such as books and pamphlets. These items end up in the Five College shared catalog and discovery layer and are contributed to WorldCat. But a user would have no way to find non-MARC archival and manuscript collections, or even uncataloged compiled collections of books from SCUA without using

UMarmot. Only patrons already familiar with SCUA know that UMarmot exists.

WHAT IT BECAME

More than 1000 records have so far been contributed to WorldCat and added to the UMass/Five College shared catalog and discovery. This creates an additional avenue for outside researchers to access UMass Amherst's unique collections. It also allows potential discovery by less-savvy users, members of the general public or undergraduates, who may not know that an internal special collections web site exists. Patrons such as these also have needs and interests in primary source materials or unique collections that they may not normally think they have access to.

Title	George Emery Papers, ca. 1900-1975.	Title and basic description	Every collection record has a "Scope & Content" note, and also a "Biographical/Historical note" if available
Author	Emery, George Edward, 1904-1975.		
Format	Collection Mixed Materials Archival Material		
Physical Descrip.	8 boxes (30 linear ft.).		
Usage Note	Contact Special Collections and University Archives, UMass Amherst Libraries, for information about access to and use of this collection.		
Scope and Content	With his passion for the circus, George Emery's papers contain material not only from his career as a circus clown in the 1920s but also from his later writings about the history of the circus, his work with the Tri-County Fair, and his long association with UMass Amherst. The collection includes correspondence with friends and family; circus toys and games; posters, photographs, and ephemera; and a library of books on circus history. Of special note are some exceptional photographs, a thick sheaf of material from the Tri-County Fair, and four remarkable scrapbooks from the internationally famous animal act Fred Kerlake's Pigs. Beginning in 1889, Fred Kerlake and his wife, Mary, traveled the world with a team of acrobatic intellectual pigs (and later donkeys) performing to packed houses. In the 1960s, Mary Kerlake gave the scrapbooks to Emery to assist in his research and writing on circus history.		
Source of Acq.	Gift of Chris Emery, July 2017.		
Biog./Hist. Note	After graduating from Massachusetts Agricultural College in 1924, George Edward "Red" Emery taught high school briefly and held a handful of other jobs before deciding to fulfill a childhood dream. Born in Marlboro, Mass., in 1904, Emery turned his love for the circus into a life touring the country as a white-face circus clown. After marrying Virginia Link, a Smith College student, in 1932, he settled down to a stable job in the Alumni Office at his alma mater, later filling in as Veterans Coordinator and as a staff member in the Student Placement Office until his retirement in 1972. Emery never left the circus behind entirely. Throughout his years in Amherst he continued to talk and write about the history of the circus and his personal experiences, and from the late 1940s through early 1960s, he used his show business connections to book talent for the Tri-County Fair. Longtime residents of Leverett, Mass., he and his wife died within a year of one another, Virginia in 1974 and George in 1975.		
Subject(s)	University of Massachusetts Amherst -- Alumni and alumnae. University of Massachusetts Amherst -- Staff.	Interesting subject headings! Clowns!	"View collection description" links to online finding aid or available description. Location and Special Collections call number also provided, both in this record and holdings record.
Subject(s)	Circus -- United States -- History. Circus performers. Clowns. Agricultural exhibitions -- Massachusetts -- History. Animal trainers. Circus animals. Swine.		
Subject(s)	Tri-County Fair.		
Other Title	George Emery Papers.		
Related Link	View collection description		
Location(s)	UM Special Coll & Archives /UM Special Collections /MS 982 LOCATION & AVAILABILITY		

All Five Colleges Call Numbers & Availability

Request Item

Emery, George Edward, 1904-1975.
George Emery Papers,
Collection

LOCATION: UM Special Collections / MS 982 [Text to phone](#)
ONLINE: [View collection description](#)

Limit by: Year Volume Location

Volume		Availability
	Library Use Only	Available

THE GOOD & THE BAD

The Five Colleges are: Amherst College, Hampshire College, Mt. Holyoke College, Smith College and the University of Massachusetts Amherst. The Five Colleges share a library catalog, off-site storage and many electronic resources. Students at any of the Five Colleges can visit any of the libraries on any campus.

The Good

To facilitate this shared educational experience, many technical services librarians have worked together over the years to standardize practices, clean up catalog data and troubleshoot discovery issues to make each campus's resources more apparent and easier to find. This project was just furthering the great work already being done.

To date:

- + Between 1500 and 2000 records have been contributed throughout the Five Colleges.
- + At UMass Amherst, approximately 120 new Name Authority Records have been contributed to the Library of Congress Name Authority File.
- + More than 50% of the Name Authority Records created by UMass Amherst represent women (<http://id.loc.gov/authorities/names/no2017121782.html>)/women's groups or members/organizations (<http://id.loc.gov/authorities/names/no2018028991.html>) from under-represented groups.
- + Two subject headings for Quaker schism groups have been added to the Library of Congress Subject Headings database.

The Bad

At UMass Amherst:

Even though it has been a great experience to work with a student employee to finish the first round of the project, it has meant that I am unable to continue making LC Name Authority Records for these collections. That is a particular value-add that was of particular interest to the librarians in Special Collections and University Archives.

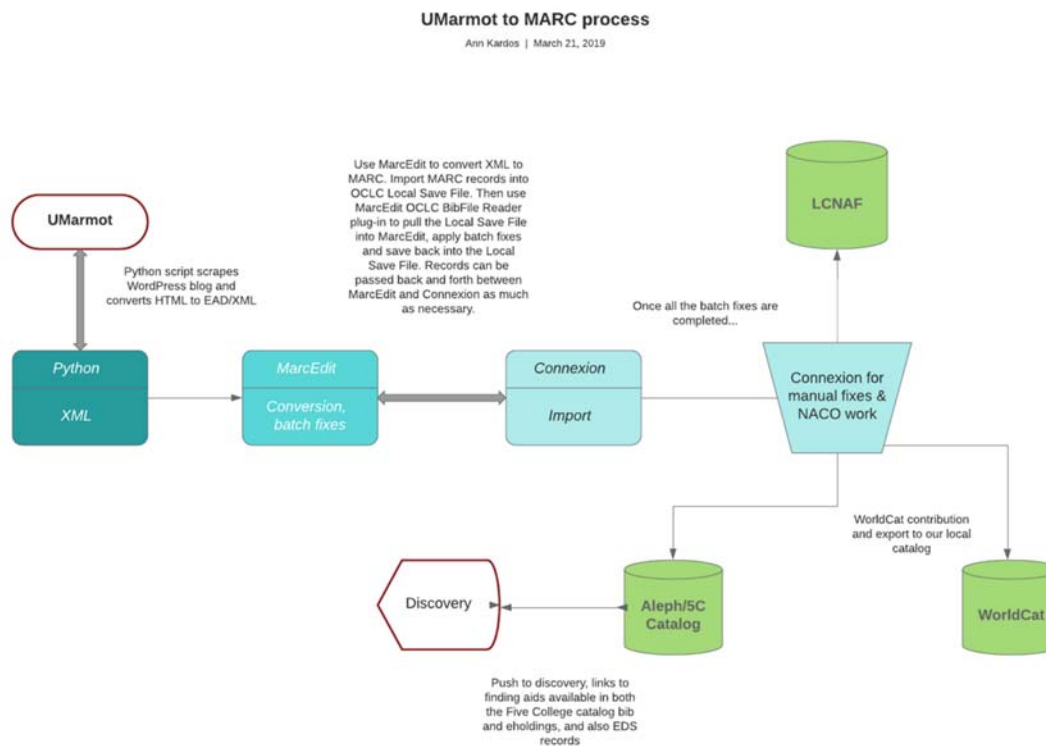
Additionally, SCUA decided not to migrate to ArchivesSpace after all. They will continue to use their UMarmot WordPress blog. They also designed and built their own tool for the rest of their needs earlier this year, in January 2019. Unfortunately, this limits the ability to make these records fully automated in the future, and puts UMass Amherst out of sync with the other Five Colleges.

For the Five Colleges:

While the ArchivesSpace to MARC Task Force is satisfied that the workflows we created are functional, there is no push-button solution that will work across our five institutions. One reason for this is the source data in ArchivesSpace. Individual local practices for collection description are varied, even over time within the same institution.

WORKFLOW

Special Collections and University Archives (SCUA) wrote a Python script to scrape and map UMarmot and generate the contents into XML. The Metadata Librarian (that's me!) receives the XML file and begins the process.



The Python scraping and subsequent XML to MARC transformation is unfortunately imperfect. For example, it maps every 6XX field to a 650 subject, and every subfield for both 6XX and 7XX fields are mapped to \$x. Also, every 1XX is mapped to 100 1_ (personal name creator) and 1XX fields are often duplicated as additional 7XX fields. As a result, even after all transformations and routine batch fixes are applied in MarcEdit, each record needs to be reviewed in Connexion to fix tags and random subfields. After completing around 900 records on my own, I created a detailed set of instructions to see if I could teach a student worker to make those kinds of judgement calls and who could understand what is essentially "MARC Lite." It's been a success, and I am now partnered with a fantastic student employee who is finishing the remaining 200 plus records from this first batch.

In addition to creating MARC records, Library of Congress Name Authority Records are created as time allows. The focus for name authority records is on collections in subject areas important to UMass Amherst, the Five Colleges and the region, in particular. Name authorities for collections representing populations from diverse and under-represented backgrounds, of both regional and wider significance, are prioritized.

UMass Amherst has regionally and nationally important collections in W.E.B. DuBois, cannabis legislation reform, pacifism, anti-nuclear activism, Quakers, gay marriage, mental health activism, back-to-the-land movements and more.

FIVE COLLEGE COLLABORATION

A Five College ArchivesSpace to MARC Task Force was convened to standardize the way each of the Five Colleges makes special collections and archival materials discoverable in the shared catalog. The goal was for the Task Force to come up with a workflow that was as automated as possible in the current environment, but with minimal effort because of a looming migration from Aleph to FOLIO. Each of the Five Colleges...except UMass Amherst...already uses ArchivesSpace, and it was rumored that UMass Amherst would soon be moving to ArchivesSpace as well.

We first sought to define a standard MARC profile for archival collections. Collection descriptions in ArchivesSpace can be quite complex and need to be simplified for MARC. Reviewing existing collection descriptions and their surrogate catalog records helped the us to identify which ArchivesSpace elements should be included in a catalog record, and which should not.

1	ArchivesSpace element	MARC element	Notes
2	If present in ArchivesSpace resource record, DO include in catalog record		
3	resource level	Leader	Static values: type = p (mixed materials); BLVL = c (collection); Ctrl = a (arc) (Full-level input by OCLC participants); Desc = i (ISBD)
4	record modified date + resource date begin/date end + language code	008	Static values: Ctry = mau, Srce = d (other)
5	resource org_code + language + finding aid descrip rules	40 Cataloging Source	Status values: \$b = ong, \$e = rda, \$e = dacs
6	resource language + source of language codes	41 Language Code	
7	resource identifier	99 Local Call #	The resource identifier is the locally assigned MS or RG number
8	resource title + date expression and/or date begin/end	245 Title Statement	
9	resource ead_location	856 Electronic Location and Access	AS exporter will place standardized language ("Finding aid online: ") in 856
10	extent number + extent type + extent container summary	300 Physical Description	
11	arrangement fileplan notes	351 Organization and Arrangement of Materials	AS exporter will not include subnotes of these types: ordered list, defined list content of an arrangement of materials note is placed in a list, the list cont
12	usersrestrict note legalstatus note	540 Terms Governing Use and Reproduction Note	
13	accessrestrict note	506 Restrictions on Access Note	
14	abstract scopecontent note	520 Summary, etc.	
15	bioghist note	545 Biog/Hist Note	
16	langhist note	546 Language Note	
17	custodhist note	561 Ownership and Custodial History	
18	accruals note	584 Accumulation and Frequency of Use Note	
19	FIRST agent with a role of "creator" of person family type	100 Main Entry/Personal Name	AS exporter takes first personal creator; all additional names go in 700
20	FIRST agent with a role of "creator" of corporate type	110 Main Entry/Corporate Name	AS exporter takes first corporate creator; all additional names go in 710

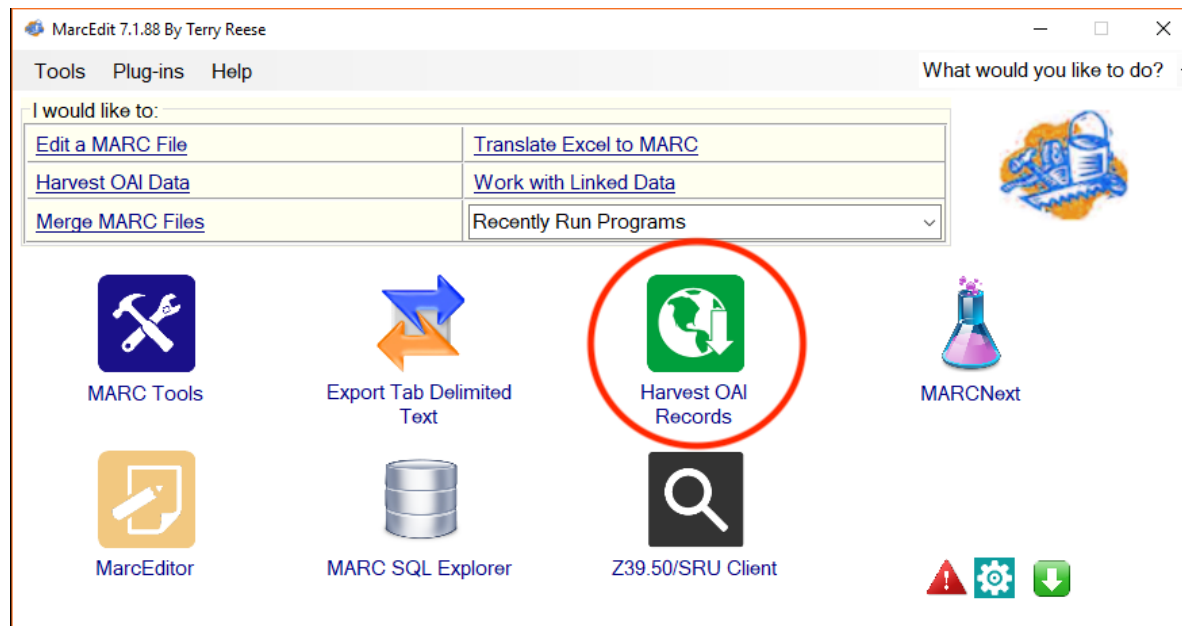
The standard ArchivesSpace MARCXML exporter was evaluated to determine whether it could be used in the workflow. We requested updated ArchivesSpace MARCXML exporter documentation and received a detailed description of how the MARCXML exporter maps ArchivesSpace elements to MARCXML. This mapping was reviewed, and we were able to indicate where the exporter created desired MARCXML, where it fell short and what should change.

We considered a number of options, including: 1) an ArchivesSpace plug-in that would allow customization of the default behavior of the exporter itself; and 2) OCLC Digital Collection Gateway, which would leverage OAI to pull data directly from ArchivesSpace on a regular schedule. Each of those options had downsides, and with the additional uncertainty around the ongoing FOLIO project, we decided that the simplest and least disruptive way to fulfill the charge was to utilize ArchivesSpace with our current existing tools and practices.

We identified approximately 20 modifications to the out-of-the-box MARCXML exporter functionality. After a test with live data, we determined that the MARCXML exporter could output the records with minimal remediation. We created two separate workflows: one uses MarcEdit, and the other is a simple XSL stylesheet. Our final workflow recommendation was submitted in March 2019.

The components of the recommended workflow include:

- 1. Retrieve records out of ArchivesSpace by harvesting with OAI-PMH or singly with the ASpace UI.



The screenshot shows a window titled "Metadata Harvester" with a sub-header "Harvest OAI Data". It features two tabs: "Home" and "Advanced Settings", with "Advanced Settings" currently selected. The form contains the following fields and controls:


- Server Address:** A dropdown menu.
- Set Name:** A text input field.
- Metadata Type:** A dropdown menu with "oai marc" selected.
- Crosswalk Path:** A text input field containing "C:\Program Files\MarcEdit 7\xslt\OAIMARC21XML.xsl" and a folder icon.
- Start:** A text input field.
- End:** A text input field.
- Debug URL:** A blue text label.

At the bottom of the window are three buttons: "OK", "Save", and "Close".

- 2. Transform the records, either with a MarcEdit Task or use XSLT.

Edit Task List
— □ ×

Task List Name:







Description:

Override Broker Action (not recommended)

Tasks:

DELETE	049	1	false	fals	^
DELETE	500	1	false	fals	
DELETE	524	1	false	fals	
DELETE	541	1	false	fals	
DELETE	544	1	false	fals	
DELETE	555	1	false	fals	
DELETE	500	1	false	fals	∨

↑
↓

Actions:    

Save
Debug
Close

-
- 3. Import records into OCLC and the Five College catalog however the individual cataloger wants--either one-by-one, or by further automation.

AUTHOR INFORMATION

Ann Kardos is a Metadata Librarian at the University of Massachusetts Amherst. She began her metadata career in museums, working with rare books, odd special collections items, 3-dimensional objects, architectural drawings and archival materials. She moved to academia in 2014 and has never looked back! Today, she splits her time between completing large scale data transformations, manipulation, analysis and batch loading (exciting!!!), original cataloging for general and special collections in multiple foreign languages (so close to her heart), learning Python and troubleshooting other people's code (eek!) and data clean-up (ok...that one can be a snoozer).

When she's not got data on the brain, she hikes mountains, rides her bike, walks her dog and drinks whiskey. She also loves Nine Inch Nails and the X-Files.

ABSTRACT

Data regarding library collections exists in many locations. While this often works for those researchers who know what they're looking for, the University of Massachusetts Amherst created a plan to convert online finding aids into MARC records to better aid search and discovery. The project gave patrons additional ways to discover unique collections, but also created an opportunity to contribute Library of Congress Name Authority Records for distinct entities or hidden individuals and groups. The project became a model for standardization of this work across the libraries of the Five Colleges, and brought together stakeholders from metadata, special collections, archives, digital collections and library IT.