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

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

Ann Kardos
University of Massachusetts Amherst
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 UMaMot (http://scua.library.umass.edu/umamot/scua/what-is-umamot/) and it also functions as their in-house searching tool.

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
UMamnot. Only patrons already familiar with SCUA know that UMamnot 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.
George Emery Papers,
Collection

**LOCATION:** UM Special Collections / MS 982  
**ONLINE:** View collection description

Limit by: Year Volume Location

Volume

<table>
<thead>
<tr>
<th>Library Use Only</th>
<th>Availability</th>
</tr>
</thead>
<tbody>
<tr>
<td>Available</td>
<td></td>
</tr>
</tbody>
</table>
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 afterall. 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 us to identify which ArchivesSpace elements should be included in a catalog record, and which should not.

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