2016

Partnering for Discoverability: Knitting Archival Finding Aids to Digitized Material Using a Low Tech Digital Content Linking Process

Liz Woolcott
Utah State University

Andrea Payant
Utah State University

Sara Skindelien
Utah State University

Follow this and additional works at: https://digitalcommons.usu.edu/lib_pubs
Part of the Library and Information Science Commons

Recommended Citation
https://digitalcommons.usu.edu/lib_pubs/269

This Article is brought to you for free and open access by the Libraries at DigitalCommons@USU. It has been accepted for inclusion in Library Faculty & Staff Publications by an authorized administrator of DigitalCommons@USU. For more information, please contact dylan.burns@usu.edu.
Partnering for Discoverability: Knitting Archival Finding Aids to Digitized Material Using a Low Tech Digital Content Linking Process

As libraries continue to ramp up digitization efforts for unique archival and special collections material, the segregation of archival finding aids from their digitized counterparts presents an accumulating discoverability problem for both patrons and library staff. For Utah State University (USU) Libraries, it became evident that a system was necessary to connect both new and legacy finding aids with their digitized content to improve use and discoverability. Following a cross-departmental workflow analysis involving the Special Collections, Cataloging and Metadata, and Digital Initiatives departments, a process was created for semi-automating the batch linking of item and folder level entries in EAD finding aids to the corresponding digitized material in CONTENTdm. In addition to the obvious benefit of linking content, this cross-departmental process also allowed for the implementation of persistent identifiers and the enhancement of finding aids using the more robust metadata that accompanies digitized material. This article will provide a detailed overview of the process, as well as describe how the three departments at USU have worked together to identify key stakeholders, develop the procedures, and address future developments.

By Liz Woolcott, Andrea Payant, Sara Skindelien

Introduction

Utah State University (USU) Library’s Special Collections and Archives and Digital Initiatives departments work together to provide scans of archival material on request to patrons in the Special Collections’ reading room. One problem that repeatedly surfaces is the request to digitize resources that are already scanned and available online. What USU Library found was an emerging pattern, where neither the patrons nor the staff members serving on the front lines (such as the Archive’s reference desk) were consistently aware of what resources were already available in digital format. Despite efforts to educate staff about local resources and also to make digital collections more discoverable by contributing them to online discovery portals such as the Mountain West Digital Library (MWDL) and the Digital Public Library of America (DPLA), USU’s digital program contained blind spots with key potential access points that remained untapped.

From both a staff and patron perspective, reasons for this issue are readily apparent. USU Special Collections staff, who assist patrons in their research, primarily utilize the online finding aids as their chief research tool.[1] [2] Finding aids are hosted online and linked from the USU Special Collections’ homepage. However, USU’s digitized archival material is housed in a separate content management system (CONTENTdm) and cannot be searched in conjunction with the finding aids. To add to this, the online finding aids rarely contain any notation that a digital component exists. Patrons and staff need to reconstruct their original searches once again within CONTENTdm in order to see if a digital version is available.

As the oft cited 4th law of Ranganathan states, the library should save the time of the reader (or, in this case, the user) (Ranganathan, 1931). Requiring duplication of searches to accommodate the cobbled together infrastructure for digital and physical formats of archival material breaches the spirit of this law. In order to bridge the gap between the digital content and the archival description, the Special Collections and Archives staff and the Digital Initiatives staff needed to work together to make digital materials accessible from multiple entry points in the simplest possible manner.

Background

USU Library formed a special working group to address workflows between Special Collections and Archives and Digital Initiatives. The working group’s name was shortened to the more tongue-in-cheek acronym of “SCAD.”

At the time that the SCAD working group was investigating workflows, metadata creation was coordinated by the Digital Discovery Librarian, but the day-to-day metadata work was split between three departments: Cataloging and Metadata Services, Digital Initiatives, and Special Collections and Archives.

At the same time, the library was also experimenting with a multi-departmental team that was cross-trained in the standards and practices of all three areas.[3] This Strategic Works and Tactics (SWAT) team was organizationally housed within Special Collections and Archives, and trained in Machine Readable Cataloging (MARC), Dublin Core, and Encoded Archival Description (EAD). Members of the SWAT team were charged with focusing on the backlogs in all three departments with the end goal of using their holistic training to provide feedback on how the library could streamline the workflows between the departments. This three-member...
SWAT team represented half of the SCAD working group with the other half composed of a Digital Initiatives assistant, an archival EAD specialist, and the Digital Discovery Librarian who chaired the group.

Over the course of a year, the group analyzed the stages of collection description at USU Libraries from the initial archival processing to the final metadata creation for the digitized surrogates. The group briefly reviewed archival management software, explored some training opportunities, and ultimately made several recommendations about practices that could be streamlined or improved. There were a lot of potential avenues to explore, so the group narrowed its focus to streamlining metadata workflows. The main product that came out of the SCAD working group effort was identifying the necessity of linking the online finding aid to the digital copy (otherwise referred to as digital content linking) and the development of a process to create those links both individually and in batches.

**Technological considerations**

Described below is the process that was developed and is currently in use by Utah State University Libraries. Before diving into that process, a few structural and technological background items should be noted.

First, the process outlined below is intended to be “low tech” wherever possible in order to reduce the technological barrier for staff and especially for the student help who do the bulk of the processing and inventory creation. Procedures requiring additional training...
or education were found to be too cumbersome to maintain long term with the naturally high turnover rates from graduating student workers. Turnover rates in front line staff are also high enough to make extensive or highly technical training unsustainable in the long run.

Second, Utah State University Library does not currently and has not historically employed any archival management software, although there is a plan for migrating to ArchivesSpace in the next two years as part of a consortia project.[4] Management of archival material is carried out through a combination of processes beginning with accessioning collections using an Access database, proceeding to physical processing and inventory creation, and ending in the construction of EAD finding aids which are stored locally on a shared drive but contributed to a consortium for hosting and public viewing.

Third, all finding aids are hand coded using the EAD 2002 XML standard.[5] There is no set process for creating EAD finding aids at USU. The process differs depending on which archival team does the processing. When developing the digital content linking process, the SCAD working group had to be cognizant of the various EAD creation styles employed by the archival teams.

Fourth, the USU Special Collections and Archives department is part of the Archives West consortium and adheres to the Best Practices Guide developed by the Content Creation and Dissemination program of that consortium.[6] The Archives West consortium is also responsible for developing the style sheet that interprets the XML structure of the finding aid and renders the final display for the viewer. The digital content linking process that was developed had to adhere to both the standards of the Best Practices Guide and also respect the capabilities of the style sheet for the consortium.

### Linking digital objects to corresponding items within EAD finding aids

**Tools Required:**
- Microsoft Excel
- Microsoft Word
- XML Editor (such as oXygen)

This is an overview of the process which utilizes robust digital collection metadata to create EAD finding aids linked to digital objects at the item or folder level using the `<dao>` tag.[7] There are three main steps to the process: First, preparing and exporting digital collection metadata. Second, editing metadata to repurpose it for digital content linking. Third, using the mail merge function in Microsoft Word to automatically create an XML format container list that can be pasted directly into the primary XML document for the finding aid.[8]

While there are undoubtedly scripts that can perform a number of the steps presented here, this process was intentionally designed to be “low tech,” requiring only familiarity with or training on spreadsheets, Microsoft Word mail merge functions, and a small bit of XML. This decision was reached by the SCAD Working Group after realizing one of the largest barriers to performing this process was the constant re-training of staff and student workers who have high turnover rates. This process was also more flexible when it came to adapting to the variety of hierarchical depths found in legacy finding aids.

#### Process Overview [9]

**Step 1. Prepare and export collection metadata from the digital asset management system in a format that can be read as a spreadsheet.**

Metadata for digital objects need to include the following fields: title, date, format, call number (or at minimum the collection, series, box, or folder information), and most importantly, the URL to access the item. In the CONTENTdm database, this is found in the Reference URL field and is automated by the system. USU Library chose to use an independent persistent identifier through the EZID service, as a best practice for long term sustainability when linking digital content, instead of relying on the Reference URL supplied by CONTENTdm. Should there be a system change, staff will not need to individually modify each of the EAD guides, catalog records, exhibits, or other haphazard places where digital content might be linked. Instead, staff can simply update the EZID database with the current location of the item and all other links will redirect correctly (DeRidder, Presnell, and Walker, 2012).

EZID is a subscription based service provided by the California Digital Library that mints DOIs and ARks as unique identifiers. To create a unique identifier, a metadata creator enters basic information about an object (name, date, creator, etc.) along with the object’s URL location.[10] Once entered, the EZID system generates both a unique identifier in the form of an ARK (e.g. ark:/85142/t4059n) and a URL which incorporates the identifier (e.g. http://n2t.net/ark:/85142/t4059n). When a user clicks on the URL generated by EZID, it redirects to the current location of the digital item. USU Libraries has traditionally embedded links to digital content in MARC records, EAD finding aids, related Dublin Core records, online exhibits, social media posts, etc. We are currently transitioning existing URLs for digital content to the EZID-generated URL in order to manage links with one service.

If an independent identifier is going to be used, it needs to be generated and recorded in the metadata for each item before the digital linking process begins. For the purpose of our process, we needed to be able to assign ARK IDs for single items or batches of items, depending on the content we were linking. For single items, the process is very straightforward – we simply filled out the EZID
online form with the title, creator, date, and URL for the item and received back the ARK and URL, which were then inputted in the metadata record. For batches of items, however, USU’s Systems Administrator developed an interface through the EZID API to load the same fields (title, creator, date, and URL) for a batch of items. The EZID service then returned a csv file with the newly minted ARKS and URLs.[11] Once created, the ARK and new URL are entered into the item level metadata by a student employee.

While this process of assigning EZID ARKs to each digital item is extremely useful in many ways, it should be noted that for the digital content linking process presented here, simply using the system-supplied URL will also work.

After ensuring the title, format, date, call number, and URL fields were correct, we exported the metadata for the collection from the digital asset management system to a tab-delimited or text file. This process varies depending on the digital asset management system used. For the bulk of the collections that USU Libraries will be linking, the metadata is downloaded from CONTENTdm.[12] This file was then opened in an Excel spreadsheet for ease in editing.

**Step 2. Edit the metadata for EAD**

We edited the spreadsheet to only include the title, format, date, call number, and URL fields necessary for the EAD finding aid and the digital content linking. Based on the complexity of the collection’s hierarchy, we inserted five empty columns at the beginning of the sheet and titled them as follows (*columns with an asterisk could vary depending on the hierarchical depth of the physical collection):

- Component Level
- Component Number
- Box*
- Folder*
- Item*

Next, we separated the call number information into series, box, folder, item, etc. There are several Excel formulas that help automate this. When all of the separated information was sorted into the appropriate columns and rows, we copied and re-pasted the columns as values, since formulas are problematic to work with in Excel. We then sorted the entries in Call Number order to reflect the hierarchical structure of the physical collection.

In the Component Level column, we indicated the component type using the approved EAD level elements.[13] In the Component Number column, we indicated the component level number (i.e., <c01> through <c12>.) However, we only listed the variable number and did not include the “c” or “c0” portion of the element. (This was added at a later stage in the process.)

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td>Component Level</td>
<td>Component Number</td>
<td>Box</td>
<td>Folder</td>
<td>Item</td>
</tr>
<tr>
<td>2</td>
<td>item</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>01:01:01</td>
</tr>
<tr>
<td>3</td>
<td>item</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>01:01:02</td>
</tr>
<tr>
<td>4</td>
<td>item</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td>01:01:03</td>
</tr>
<tr>
<td>5</td>
<td>item</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>01:02:01</td>
</tr>
<tr>
<td>6</td>
<td>item</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>01:02:02</td>
</tr>
<tr>
<td>7</td>
<td>item</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>01:02:03</td>
</tr>
<tr>
<td>8</td>
<td>item</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td>1</td>
<td>01:03:01</td>
</tr>
<tr>
<td>9</td>
<td>item</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td>2</td>
<td>01:03:02</td>
</tr>
<tr>
<td>10</td>
<td>item</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td>3</td>
<td>01:03:03</td>
</tr>
</tbody>
</table>

**Figure 2.** Adding Component Level and Component Number Information (Note: this particular example has no hierarchy)

Next, we added a column at the end with the header: “Resource Label.” Underneath that column header, we entered the text “Click to Access” for each row in the column. This is the hyperlinked text that is displayed to the user.
The final spreadsheet had the following columns (*columns with an asterisk varied depending on the physical collection’s hierarchical depth)

- Component Level
- Component Number
- Box*
- Folder*
- Item*
- Title
- Format
- Date
- URL
- Resource Label

Finally, we added additional rows as necessary to create level information for series, boxes, or folders that are not digital objects, but would serve as headers in the EAD finding aid. (see rows 2 and 3 in Figure 3 below for an example.)

**Figure 3.** Example of a final spreadsheet (using collection data as seen in Figure 1 – Note: this collection has a hierarchy). Click image for larger view.

**Step 3. Use the Mail Merge function to create a new XML container list with links to digital content**

Once finished making the necessary edits to the spreadsheet, the next step was to utilize the Mail Merge function in Microsoft Word to create a new XML container list for the finding aid embedded with the new links to digital content.

To begin, we opened a new Word document and inserted an XML template like the one listed below. This template represented the XML coding needed for a single item in a collection. We were sure to include the digital archival object and any attribute tagging necessary for the content linking to operate effectively.

**Template for use with xlink namespace:**

```xml
<?xml version="1.0"?>
<container>
  <component>
    <title>Component Title</title>
    <level>Component Level</level>
    <number>Component Number</number>
    <box>Box</box>
    <folder>Folder</folder>
    <item>Item</item>
    <title>Title</title>
    <format>Format</format>
    <date>Date</date>
    <url>URL</url>
    <resourceLabel>Resource Label</resourceLabel>
  </component>
  <!-- Repeat for each item -->
</container>
```
The Component_Number, Component_Level, Box, Title, ARK_URL, Title, Format, and Date elements in the templates above represent where the data from each column in the spreadsheet will be placed. As each row in the spreadsheet represents a new digital item, it creates a new <c0> unit for each row.

Using the mail merge function in Word, we assigned columns from the spreadsheet to the corresponding EAD elements in the XML template and ran the merge. This produced a new Word document. The XML coding for each of the items in the collection was visible with information inserted from the spreadsheet. One entry was displayed on each page of the document. We then removed the extra blank space between entries by using the Find and Replace function in Word (using ^b to find blanks, and replace with ^|^|) to insert two empty lines or manual line breaks between the entries.

The end result looks like this:
Once finished, we copied the container list in Word and pasted it into the <dsc> section of the XML file for the collection’s finding aid. Depending on the collection’s need, we might overlay the entirety of the previous <dsc> content or just insert individual links (or smaller groupings of links.) We then performed quality control on the XML.

NOTE: If the digital collection is not a 1-to-1 relationship with the physical collection, some edits may need to be made. It may be advisable to only copy and paste specific sections of the newly minted XML. If the digital content only reflects random individual items in a collection, the process would need to be adjusted to copy and paste single items at a time.

Figure 4. A completed XML container list in word.
Figure 5. The finding aid complete with links to digital objects in Archives West (Note: folder information is not seen here because it was added in the above procedures for demonstration purposes only. For this collection, “1972.001.003” is a single item number and not a box or folder number.)

Notes on the Benefits and Considerations of Digital Linking

This batch process can be used in two ways to provide digital content links: to formulate a brand new XML container list for an EAD finding aid or to update an existing container list. With regard to updating an existing EAD container list, this process is most accommodating for digital collections that share a 1-to-1 relationship with the physical collection, where all items in the physical collection have been digitized and are housed in one single digital collection. Digital collections that contain more than one physical collection will be more problematic and require additional steps to sort. Likewise, digital collections with content that comprise only a portion of a physical collection, instead of all of it, will also require additional steps.

While the overall benefit of creating digital links between a finding aid and digital images may appear quite obvious, there were several elusive benefits and considerations we uncovered as we developed this process.

Robust finding aid inventories

Prior to standardization through the advent of DACS (Describing Archives: A Content Standard) or EAD (Encoded Archival Description), finding aids were as unique as the people arranging and writing them. These legacy finding aids were often composed with seasoned researchers as their target audience—researchers who journeyed into special collections to view materials first hand. The digital age has changed all that and the expectation to have more materials in a digital format has increased. By linking individual items from an archival finding aid to a digital object, researchers are not only provided with the organizational context of archival materials found in a finding aid, but also a visual object that otherwise may not be easily accessible. Described by Zhang and Mauney (2013) as the parallel model, digital linking on the item, folder, or box-level allows the archival materials to be presented in two distinct ways without forsaking one format for another. The finding aid preserves the contextual information researchers need to understand provenance and historical information about the collection, while the metadata associated with the digital file provides more detailed, granular, and controlled descriptions that can be faceted within and across collections to tie disparate items together. This process allows us to retain the benefits inherent in both of the parallel systems we were using and also unites them together at the point of discovery. Without any training at all, front line staff or users who have stumbled upon the EAD guide can access the digitized version of an item without having to re-do their searches in a separate content management system (assuming that they know that separate system exists in the first place.)

Due to the nature of descriptive metadata creation, titles for digital assets are often more detailed than what is recorded in the EAD guide. When digital asset metadata is transformed into an EAD environment, these more detailed titles are transferred with them and provide important keywords to improve discoverability. This process can serve as a complement to the archival “More product, less process” or “MPLP” philosophy. (Greene and Meissner, 2005) Material can be inventoried in an EAD guide using the MPLP process, but then be iteratively upgraded using metadata descriptive practices when a collection is deemed worthy of digitization.
New accessioning opportunity

The digital content linking process also allows for the “quick” creation of finding aids for newly acquired or born-digital collections. USU Libraries experimented with taking newly created born-digital objects from an oral history project conducted by our Fife Folklore Archive and accessioned them using the digital content linking process. In effect, the digital files were uploaded to CONTENTdm and described by metadata specialists rather than an archival processing team. Once completed, the metadata was exported, edited as described above, merged into an XML template using the mail merge process and then inserted into the portion of a brand new finding aid. The curator then created the “front matter” (or background, scope and content, and other information regarding the provenance of the collection.) This streamlined process provided a new potential model for accessioning collections where metadata specialists and curators worked in concert to bring both the digital collection and the finding aid online simultaneously. Of course, this will not work for all newly acquired collections, particularly large ones or collections where patron demand is unknown, but it does provide a possible new model for smaller collections where high patron demand is anticipated.

Feedback

In May of 2016, the authors presented this process at the Conference of Inter-Mountain Archivists (CIMA) as part of a panel on re-purposing archival metadata. Immediately following the presentation, we received multiple requests for the step-by-step process instructions as many institutions were interested in how we maintained a low technological bar by using the programs and software they also use.

Library and archival staff have also been quite receptive to the process of inserting digital links between finding aids and digital content. They appreciated the fact that student training is very minimal and the tutorial developed for this process allows for quick training sessions, particularly since the student employee turnover rate is high in the academic world. As a result of this staff support, we have begun to collaborate with other departments outside of the archives to train more student workers on the process.

To manage the “backlog” of finding aids needing digital linking, we organized our finding aids into three phases according to the level of complexity. Phase I is the direct one-to-one matches between the finding aid and the digital collection (where a single archival collection has been digitized and put online as a single digital collection); Phase II includes all of the digital collections that contain multiple finding aids (where a single digital collection has items from more than one physical archival collection); and Phase III is comprised of digital collections that do not have a corresponding finding aid but are excellent candidates for the reversal process of creating a finding aid inventory from the digital collection metadata while simultaneously inserting digital links (described in the “New accessioning opportunity” listed above).

Conclusion

Digital content linking provides the missing step between EAD finding aids and digital content that is housed outside of the EAD environment. It allows users and library staff to know which items in a collection have been digitized and brings them directly to the digital objects. Providing cross linking of finding aids to digital content and vice versa creates cyclical access to accommodate users from any discovery point. Creating digital content links one by one may be appropriate for situations where only select items in a collection are digitized, but for collections that have large portions of content digitized, bulk procedures are the most efficient means for creating access points.

The procedures outlined in this article can prove to be beneficial by adding valuable information to the EAD finding aid. However, changes may affect the look and feel of the finding aid and should be discussed with archivists and curators before proceeding. Additional discussions about the appropriate level to digitize and describe an item are also imperative. While these procedures were developed for archival collections with existing finding aids, they were also found to be easily adapted to help streamline workflows for born-digital content.

Notes

[1] Also called inventories, finding aids, finding guides, registers, etc. In this paper, we will refer to them as finding aids for easy reference. They are often marked up in Encoded Archival Description (EAD) XML. When discussing the XML, we will refer to it as an EAD finding aid.

[2] It should also be noted that patrons are also accessing the finding aids when not in the reading room, as well. Analytics for finding aid access shows that they are receiving a robust amount of traffic off campus.

[3] The SWAT team pilot project was reported on at the American Library Association Annual Conference in 2014. To see more information on the structure and planning of it, please visit: http://digitalcommons.usu.edu/lib_present/61/
This migration will change the procedures outlined below in several substantive ways, but we anticipate that much of the process can still be retained after adapting it to the Archives Space requirements.

In concert with partners in the Archives West consortium, USU Library will not be migrating to the new EAD3 standard until after 2018.

The Best Practices guide is available here: https://www.orbiscascade.org/file_viewer.php?id=2923

To see the specifications for the tag in EAD 2002, please visit: https://www.loc.gov/ead/tglib/elements/dao.html

The mail merge process was not created by Utah State University, but was shared by the Utah State Archives with USU Libraries in 2008. It can be found here: http://archives.utah.gov/research/inventories/ead.html

To see the step-by-step instructions for this process, please visit the USU Cataloging and Metadata Services blog: https://usucataloging.wordpress.com/2016/09/26/digital-content-linking-workflow/

You can find out more about EZID at their website: http://ezid.cdlib.org/

This code is currently not publicly available. USU Libraries is investigating the possibility of opening the source code in the future.

This process will vary depending on the digital asset management system. For CONTENTdm users, the process is outlined here: https://www.oclc.org/support/services/contentdm/help/collection-admin-help/exporting-metadata.en.html

EAD tag library: https://www.loc.gov/ead/tglib/att_gen.html


About the Authors

Liz Woolcott
liz.woolcott@usu.edu
Liz Woolcott is the Head of Cataloging and Metadata Services for Utah State University Libraries. She has worked in cataloging and metadata coordination for the last 12 years and is the co-founder of the Library Workflow Exchange.

Andrea Payant
andrea.payant@usu.edu
Andrea Payant is the Metadata Coordinator for Utah State University Libraries and has worked for USU for 10 years. Her most recent work has focused on data management and consulting for library staff as well as students and faculty campus-wide.

Sara Skindelien
sara.skindelien@usu.edu
Sara Skindelien is the Special Collections and Archives Library Assistant and specialist in EAD for Utah State University Libraries. She holds an MA in Public History from the University of Northern Iowa.