Documenting a Move Using Archival Description: Tools for Bridging the Gaps Between Physical and Intellectual Control

Bolton Doub
University of Southern California, bdoub@usc.edu

Follow this and additional works at: https://digitalcommons.usu.edu/westernarchives

Part of the Archival Science Commons

Recommended Citation
Doub, Bolton (2022) "Documenting a Move Using Archival Description: Tools for Bridging the Gaps Between Physical and Intellectual Control," Journal of Western Archives: Vol. 13: Iss. 1, Article 11. Available at: https://digitalcommons.usu.edu/westernarchives/vol13/iss1/11

This Case Study is brought to you for free and open access by the Journals at DigitalCommons@USU. It has been accepted for inclusion in Journal of Western Archives by an authorized administrator of DigitalCommons@USU. For more information, please contact digitalcommons@usu.edu.
Documenting a Move Using Archival Description: Tools for Bridging the Gaps Between Physical and Intellectual Control

Bolton Doub

ABSTRACT

Following the move of approximately 40,000 linear feet of archival material between offsite storage facilities, the University of Southern California (USC) Libraries began a project to document these holdings’ new locations using ArchivesSpace. This case study explores a combination of tools—including the ArchivesSpace API, Python scripts, OpenRefine, and spreadsheet applications—that the USC Libraries used to batch-edit and create container data in ArchivesSpace following the move. The paper discusses the challenges and shortcomings of these tools for editing particular forms of legacy data entered into USC’s instance of ArchivesSpace long before the move. When the creators of this past description prioritized the work of establishing intellectual control (describing the informational content of archival resources) using methods that neglected descriptive prerequisites for the future maintenance of physical control (tracking the physical locations of archival holdings), the tools outlined in this paper were less effective in editing that legacy data.

Introduction

Beginning in 2019, the University of Southern California (USC) Libraries started a project to move approximately 40,000 linear feet of archival material out of an offsite storage building known as “East Library.” This case study focuses on the work directly following the physical relocation of the material: the work of updating the tens of thousands of container records describing each box in USC’s archives management application, ArchivesSpace.

The project to update all of these container records required the use of various tools, including Python and Java scripts, the ArchivesSpace API, OpenRefine, and Excel. However, the efficacy of these tools depends on a number of prerequisites relating to how archivists initially described these boxes and their contents. This case study will present a combination of tools that the USC Libraries used to batch-edit archival description. The study will also explore the unique sets of challenges associated with batch-editing different forms and levels of archival description. In
examining these methods, the case study highlights key disconnects between the labor involved in establishing intellectual control (describing the informational content of archival resources) versus maintaining physical control (tracking physical locations of archival holdings).

Literature Review

A significant amount of the literature on moving archival collections has focused on the planning and preparation phases of relocations, which is related but also somewhat separate from the post-move emphasis of this case study. Pam Hackbart-Dean, Leah Agne, and Julie Mosbo authored a paper on the planning and execution of moving the Southern Illinois University Carbondale’s (SIUC) Special Collections holdings to a new storage facility in 2010. While the majority of the article on SIUC’s move outlined the preparation phases of their relocation project, Hackbart-Dean et al. briefly note the work of creating new, post-move shelf lists to check against pre-move shelf lists documented in spreadsheet files. The article does not specify tools used—beyond spreadsheets—in creating these shelf lists or if SIUC copied or exported the new locations from these shelf lists into a data management application, such as ArchivesSpace or Archivists’ Toolkit.

More recently, Claudio Gómez, Cristian Becker, and Leslie Azócar wrote a similar case study on the Museo Nacional de Historia Natural’s (MNHN) offsite collection storage project completed in 2019. In both the SIUC and MNHN case studies, the authors describe comprehensive assessment and description projects undertaken as part of the pre-move phases of each project. SIUC and MNHN invested considerable time and resources to update descriptions, create new shelf lists, and in some cases rehouse the holdings that were moving to new storage spaces. The MNHN rolled out their pre-move “Collections Management Project” over nine years to prepare their materials for relocation. In contrast, the USC Libraries had less than a year’s notice to prepare for the relocation of the 40,000 linear feet of archival collections moving out of East Library. In the case of USC’s move, staff and contractors had to rely heavily on existing collections documentation paired with incomplete post-move shelf lists to track the new locations of USC’s relocated holdings.

One paper about moving archival materials that gets relatively specific about tracking locations using both pre- and post-move collections documentation systems is Mary LaFogg and Christine Weideman’s case study concerning the move of Yale University Library’s special collections holdings to offsite storage. LaFogg and


Weideman go beyond discussions of maintaining physical control over relocated holdings and address the aspects of intellectual control involved in any move of special collections materials. Specifically, LaFogg and Weideman describe their use of a custom script, written in-house, to upload each container’s accession and barcode information from Yale University Library’s local master record for container locations to its larger online catalog system, Orbis. LaFogg and Weideman describe how Yale University Library used its collections move as an opportunity to migrate and centralize various types of collection data, ending up with unique container-level records for all of its special collections holdings in the library’s online catalog.3 The USC Libraries is taking a similar approach with the move of its archival collections, using its recent relocation projects as opportunities to synchronize and improve existing archival description across multiple systems.

East Library

The offsite storage building where the USC Libraries had been storing the 40,000 linear feet of archival material prior to the 2019-2020 move is known as East Library (Fig. 1). The USC Libraries started storing archival material in East Library in approximately 1984. Over the subsequent decades, East Library, which was built in 1927 as an industrial manufacturing facility, fell into a state of disrepair. Most of the building, with the exception of a few rooms, did not have heating, ventilation, or air conditioning during the time that it was used as storage for the USC Libraries. East Library housed a mix of processed and unprocessed archival collections, along with archival supplies, old furniture, and non-circulating books. USC Libraries staff and faculty regularly paged material from East Library to Doheny Memorial Library to fulfill researchers’ requests—mostly for material held by the Special Collections, Cinematic Arts Library, and East Asian Library units of the USC Libraries.

Among the staff and faculty of the USC Libraries, East Library became infamous for its substandard storage conditions. However, due to the cost of alternative storage solutions in central Los Angeles and the 2014 designation of the East Library building as a California Historic Resource (as an example of simplified Beaux Arts and Art Deco architecture), USC continually stalled and postponed the Libraries’ calls to move the collections out of the building. Finally, after the Libraries received support in late 2019 from the USC Office of the Provost to relocate, the Libraries were charged with moving all collection materials out of East Library by the end of 2020.

Documenting the Move from East Library to Grand Archives

The USC Libraries contracted with an independent library services company to manage much of the physical relocation of collection materials from East Library to Grand Archives. The destination for these materials, Grand Archives, is a large area of shelving in a USC-owned building—located on Grand Avenue—with much better storage conditions, including HVAC, compared to East Library. The library services company started on the larger archival collections that were relatively well documented. After the contractors moved a batch of these larger collections, they would share an Excel spreadsheet listing the new Grand Archives shelving coordinates for each collection. Each spreadsheet (Fig. 2) would have a row for each box with columns for box number, collection number, collection title, and three
numbers separated by slashes indicating the Grand Archives range, section, and shelf coordinates for each box. The USC Libraries provided an example shelf list spreadsheet to the independent contractor, which the contractors used as a template.

<table>
<thead>
<tr>
<th>Container indicator</th>
<th>notes</th>
<th>Collection no.</th>
<th>Collection title</th>
<th>Range/Section/Shelf</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td>159</td>
<td>Lawrence Lipton papers</td>
<td>120/11/9</td>
</tr>
<tr>
<td>2</td>
<td></td>
<td>159</td>
<td>Lawrence Lipton papers</td>
<td>120/12/1</td>
</tr>
<tr>
<td>3</td>
<td></td>
<td>159</td>
<td>Lawrence Lipton papers</td>
<td>120/12/2</td>
</tr>
<tr>
<td>4</td>
<td></td>
<td>159</td>
<td>Lawrence Lipton papers</td>
<td>120/12/3</td>
</tr>
<tr>
<td>5</td>
<td></td>
<td>159</td>
<td>Lawrence Lipton papers</td>
<td>120/12/4</td>
</tr>
<tr>
<td>6</td>
<td></td>
<td>159</td>
<td>Lawrence Lipton papers</td>
<td>120/12/5</td>
</tr>
<tr>
<td>7</td>
<td></td>
<td>159</td>
<td>Lawrence Lipton papers</td>
<td>120/12/6</td>
</tr>
<tr>
<td>7A</td>
<td></td>
<td>159</td>
<td>Lawrence Lipton papers</td>
<td>131/9/6</td>
</tr>
<tr>
<td>8</td>
<td></td>
<td>159</td>
<td>Lawrence Lipton papers</td>
<td>120/12/7</td>
</tr>
<tr>
<td>9</td>
<td></td>
<td>159</td>
<td>Lawrence Lipton papers</td>
<td>120/12/8</td>
</tr>
<tr>
<td>9A</td>
<td></td>
<td>159</td>
<td>Lawrence Lipton papers</td>
<td>131/9/7</td>
</tr>
<tr>
<td>10</td>
<td></td>
<td>159</td>
<td>Lawrence Lipton papers</td>
<td>120/12/9</td>
</tr>
<tr>
<td>11</td>
<td></td>
<td>159</td>
<td>Lawrence Lipton papers</td>
<td>120/13/1</td>
</tr>
<tr>
<td>12</td>
<td></td>
<td>159</td>
<td>Lawrence Lipton papers</td>
<td>120/13/2</td>
</tr>
<tr>
<td>13</td>
<td></td>
<td>159</td>
<td>Lawrence Lipton papers</td>
<td>120/13/3</td>
</tr>
<tr>
<td>14</td>
<td></td>
<td>159</td>
<td>Lawrence Lipton papers</td>
<td>120/13/4</td>
</tr>
<tr>
<td>15</td>
<td></td>
<td>159</td>
<td>Lawrence Lipton papers</td>
<td>120/13/5</td>
</tr>
</tbody>
</table>

Figure 2. Part of a shelf list spreadsheet for the Lawrence Lipton papers created by the independent contractor after moving the collection

The large, better-documented collections that moved out of East Library first typically already had a combination of description in ArchivesSpace and Alma—USC Libraries’ two main record systems for managing archival collections. In ArchivesSpace, each of these priority collections had a published record with some form of container and location information either linked to the collection-level record or more granular box- and/or folder-level records. Before the move, these ArchivesSpace container and location records documented the East Library locations for each box, including the building and room number, along with the shelving range, section, and shelf coordinates. Many, but not all, of these boxes also had a barcode affixed to the outside of each box, but these box barcodes were only documented in Alma, USC’s integrated library system (ILS). USC Libraries staff had not been entering barcodes into ArchivesSpace prior to this move.

Between Alma and ArchivesSpace, each system had its strengths and limitations for documenting these boxes (and their relocation). At USC, ArchivesSpace is the system that holds the most complete form of each archival collection’s finding aid, with meaningful links between the boxes themselves and the box- and folder-level description structured in published finding aids. Researchers use these finding aids to
request specific parts of collections—and the container locations in ArchivesSpace are used in each box’s retrieval. ArchivesSpace also has a greater capacity for tracking detailed location information for each box, with the granular range, section, and shelf coordinates already in use for tracking East Library locations.

USC’s instance of Alma holds much less description relating to archival collections compared to ArchivesSpace. In the case of USC’s archival description, Alma only holds collection-level records without the more detailed box- and folder-level description from the finding aids in ArchivesSpace. Alma location codes are also much less specific for archival collections at USC—omitting the detailed shelving coordinates that the Libraries track in ArchivesSpace. Some of USC’s Alma location codes for archival materials did not get any more specific than the building that housed the material. For example, “spe-eassto” was the Alma location code for any Special Collections archival materials stored in East Library, regardless of floor, room, or shelving coordinates. The one strength that Alma had over ArchivesSpace was that most of the box barcodes for these archival collections had been entered in Alma as “Item” records. If USC’s move out of East Library had a longer timeframe, allowing for planning phases similar to the case studies referenced in this paper’s literature review, the Libraries may have been able to copy all of the barcodes from Alma to ArchivesSpace prior to the boxes’ relocation. However, given the short timeframe and the limited number of staff trained in using ArchivesSpace, the Libraries decided to prioritize tracking each box’s new location in ArchivesSpace without the use of box barcodes during the initial phases of documenting the move.

Staffing the Project to Document the East Library Move

The USC Libraries assigned a minimal level of staffing toward documenting the move of archival collections out of East Library. Most of the work outlined in this paper was accomplished by one person: the Archival Projects Librarian in the Special Collections unit of the Libraries. The core responsibilities of the Archival Projects Librarian (also the author of this paper) at the time of the East Library move included accessioning new archival acquisitions, processing and deaccessioning existing archival holdings, providing reference services to Special Collections users, and helping with the unit’s outreach and instruction efforts. The Archival Projects Librarian’s capacity to document the move out of East Library largely depended on the timing of the COVID-19 pandemic and USC’s work from home mandate during 2020.

The independent contractor moved the first batch of archival collections (comprising approximately 20,000 linear feet) out of East Library during the months of January, February, and early March of 2020. USC’s work from home period began on March 16, 2020. If the resulting shift in the Archival Projects Librarian’s responsibilities had not occurred at this time, the timeline for documenting the move out of East Library would have been significantly longer. With the work from home shift, the Archival Projects Librarian was able to use the tools outlined in this paper to update the locations of over 8,000 existing container records and create over 11,000
new container records in ArchivesSpace for the relocated collections during spring and early summer of 2020.

Part of the time that the Archival Projects Librarian spent on this project consisted of researching and testing the tools described in this paper. Before the move out of East Library, no one in the USC Libraries Special Collections unit had accessed and used the ArchivesSpace API. Zahid Rafique, MIS Director for the USC Digital Library, had used the ArchivesSpace API before to bulk-create and bulk-edit digital object records linking to the Digital Library. Zahid was able to help with one of the scripts described in this paper, but since the move documentation project was not directly related to the USC Digital Library, the time that Zahid could commit to the project was limited. For the other tools described in this paper, the Archival Projects Librarian relied on the ArchivesSpace user community distributed among various institutions, particularly: Mark Custer, Archivist and Metadata Coordinator at Yale University Library; Vakil Smallen, International Brotherhood of Teamsters Labor History Archivist at George Washington University; and William David (Dave) Mayo, Senior Digital Library Software Engineer at Harvard Library. The Archival Projects Librarian used various channels to reach this community and its documentation, including the ArchivesSpace Users Group listserv, the Archivists Working with Archival Data Slack Workspace, and existing documentation posted on GitHub, YouTube, and the websites for the Society of American Archivists and Lyrasis.

Updating Locations in ArchivesSpace—First Steps

Once USC Libraries decided that ArchivesSpace would be the main system for tracking the archival collections’ new locations, the first step was to come up with a plan of action. The ArchivesSpace user community was a useful source of information in drafting this plan. In January of 2020, the Archival Projects Librarian posted a call for help to the ArchivesSpace Users Group listserv titled “Tools for batch updating locations in ArchivesSpace during a collections move.”4 This email introduced the relocation project and described how most of the boxes being moved already had locations in ArchivesSpace. One additional challenge disclosed in the user group post was that the capacity of each destination shelf in the new building would not necessarily match the capacity of the corresponding shelving in the old offsite storage building. In many cases, a set of boxes that shared the same shelf in the old facility would be divided into different shelves in the new facility. Due to this latter challenge, the Libraries would not be able to just edit the existing East Library location records in ArchivesSpace (with their existing links to box records) to reflect each new destination shelf in Grand Archives.

Mark Custer, Archivist and Metadata Coordinator at Yale University Library, responded to USC’s call for help to the ArchivesSpace Users Group with a helpful

Mark’s suggested course of action boils down to four main steps. First, create new locations for the new offsite facility shelving and create a corresponding spreadsheet that lists each new location record’s ArchivesSpace URI (uniform resource identifier). Second, use the ArchivesSpace API to extract all of the existing container records (generally, a record for each box) from ArchivesSpace. The container data from this second step would need to include each container record’s ArchivesSpace URI, which is not included in the reports that users can run from the ArchivesSpace staff interface. Third, create a spreadsheet with two columns: one column listing each moved container record’s URI and the second column listing the corresponding new location URI for each container. Last, use the ArchivesSpace API and a Java or Python script to re-associate the container records with their new locations based on the spreadsheet from the previous step. After all of these steps, each box record in ArchivesSpace will be re-linked to its new location.

The first step in the plan above, creating new location records for the destination shelving, was the simplest step in this process. Creating new location records does not require the use of Python scripts or the ArchivesSpace API. ArchivesSpace users can use the ArchivesSpace staff interface to batch-create new location records using either the “Create Batch Locations” tool or the “Import Data—Location CSV” function. The Create Batch Locations tool uses ranges of shelving coordinates (e.g., range, section, and shelf) to create up to 1,000 new location records at a time. The USC Libraries created these new location records with a linked “Location Profile,” which specifies the dimensions of each shelf, for subsequent shelf space planning. Creating new location records with linked location profiles is also a prerequisite for ArchivesSpace’s “Space Calculator” to function, which users can run to search for available shelf space when shelving new boxes.

As previously mentioned, each new location record’s ArchivesSpace URI will become an essential component of the batch-update-locations script detailed in the next sections of this paper. Due to the importance of the location URIs, USC Libraries batch-created its new location records in order, starting with the lowest (numerically-speaking) range number. This creation order made it easier for the Libraries to list all of the URIs for these new location records. After all of the new location records for the Grand Archives shelves were created, the archivist used the “Download CSV” function from the “browse locations” page in the ArchivesSpace staff interface to get a human-readable list of USC’s ArchivesSpace locations in a spreadsheet. The archivist then used Excel to filter out all of the non-Grand Archives location records from the spreadsheet. In order to add the location records’ URIs to this spreadsheet, the

archivist opened the ArchivesSpace staff interface and navigated to the record of the first location created during the “Create Batch Locations” process: [Range: 103, Section: 1, Shelf: 1]. Each location record’s page in the staff interface displays the record’s URI (for the 103/1/1 location, the URI is: “/locations/20080”). Each subsequent location URI followed in order, with Range: 103, Section: 1, Shelf: 2 having URI “/locations/20081.” Since all of these location records were batch-created in order, mapping each of the new location records with its unique URI in a spreadsheet (Fig. 3) was relatively simple. In a scenario where the location records were not created in a usable order, the ArchivesSpace administrator(s) would be able to use the ArchivesSpace API and a script to get the location URIs, but the USC Libraries did not need to use such a method in this case.

Figure 3. Spreadsheet of Grand Archives location URIs.

Using the ArchivesSpace API to Gather Container URIs

The first step in this case study involving the use of the ArchivesSpace API was gathering all of the container metadata from ArchivesSpace. The USC Libraries used a slightly altered version of a Python script posted by Vakil Smallen at George Washington University titled “Get_Top_Containers from One Collection-
Anonymized. One of the reasons why the USC Libraries used this script was because Vakil featured the script in a webinar and corresponding blog post with detailed step-by-step instructions for accessing the ArchivesSpace API and running the `Get_Top_Containers` script (among other scripts also featured in the webinar). As previously mentioned, no one in the USC Libraries Special Collections unit had worked with the ArchivesSpace API directly prior to this project, so Vakil’s instructions were very useful. The one small change that the USC Libraries made to the `Get_Top_Containers` script was to pull data for all of the container records in an ArchiveSpace repository, instead of limiting the records to a single collection. The USC Libraries accomplished this change by removing the line that reads “if XXXXX in collection” from the script.

The USC Libraries used Jupyter Notebook, the web-based application that Vakil recommends in his webinar, to run the `Get_Top_Containers` script and connect to USC’s ArchivesSpace API. The `Get_Top_Containers` script includes a line where the user lists the URL address to a specific instance of the ArchivesSpace API. The USC Libraries uses Atlas Systems Support to host its instance of ArchivesSpace, so the archivist running the script had to contact Atlas Systems Support to get the URL for USC’s ArchivesSpace API. After connecting to USC’s ArchivesSpace API and running the `Get_Top_Containers` script for the Special Collections repository, USC had a JSON file documenting approximately 25,000 container records, including each container record’s unique ArchivesSpace URI. The USC Libraries used the open-source application OpenRefine to convert the JSON file to an Excel spreadsheet file (Fig. 4). However, recent versions of Excel can also perform the JSON to Excel conversion directly via the software’s “Get & Transform Data” functions.

Matching Container and Location URIs

At this point in the process, the USC Libraries had access to three different sets of data: (i) the shelf list spreadsheets created by the independent contractor listing each box’s new shelving coordinates (Fig. 2), (ii) the container records (including each container record’s ArchivesSpace URI) from running the `Get_Top_Containers` script (Fig. 4), and (iii) the spreadsheet of Grand Archives location URIs created after using the “Create Batch Locations” tool in ArchivesSpace (Fig. 3). Combining all three of
these data sets was the most time-consuming part of the post-move documentation process. The end goal of this step in the process was to create spreadsheets with two columns for all of the boxes moved from East Library to Grand Archives. The first column would list all of the ArchivesSpace container URIs and the second column would list the corresponding location URI for each container’s new Grand Archives shelf.

The first step in compiling the container/new location URI spreadsheets was to combine the shelf lists from the independent contractor with the container records from ArchivesSpace to ensure that each set of data corresponded with each other in terms of the number of boxes for each collection. This combined data also paired each container URI with its new (human-readable) Grand Archives shelving coordinates. The independent contractor recorded the new shelving coordinates as three numbers separated by slashes (e.g., “103/1/1” for Range: 103, Section: 1, Shelf: 1). The USC Libraries would compile these spreadsheets for a few large collections at a time. This step in the process moved much faster when the container information from ArchivesSpace was accurate and complete, which was not the case for some of the collections that moved.

Once the USC Libraries had a combined spreadsheet for a few large collections where the box information from the independent contractor and ArchivesSpace matched, the next step was to add the corresponding ArchivesSpace location URI for each human-readable combination of shelving coordinates. The Libraries used

![Spreadsheet of ArchivesSpace container records from Get_Top_Containers script.](image)
another Python script described in Vakil Smallen’s webinar for this step. In this case, the script that Vakil titled “Merge Two CSV Files with A Common Data Point” was used to bring the ArchivesSpace location URI into the combined spreadsheet from the previous step. The two spreadsheets being combined in this step consist of: (i) the spreadsheet listing each new Grand Archives location record with its ArchivesSpace URI (Fig. 3) and (ii) the spreadsheet from the previous step that combines the box information from ArchivesSpace and the shelf lists created by the independent contractor that moved the boxes (Fig. 5).

The common data point on which both of the spreadsheets would merge was the new Grand Archives shelving location (the three shelving coordinates separated by slashes). In both spreadsheets, the column containing the shelving coordinates was labeled “ShortenedShelf” in the header row. Following Vakil’s instructions, the Archival Projects Librarian used Jupyter Notebook again to run the script to merge the spreadsheets on the “ShortenedShelf” column. After running the merge script, the Archival Projects Librarian had a copy of the spreadsheet pictured in Figure 5, but with one additional column listing the ArchivesSpace location URI for each “ShortenedShelf.”

![Figure 5. Spreadsheet combining container data from ArchivesSpace (columns A and B) and the shelf lists created by the independent contractor that moved the collections (columns C and E).](https://digitalcommons.usu.edu/westernarchives/vol13/iss1/11)

There are other methods for merging spreadsheets using a common datapoint that do not require the use of a Python script. After the Archival Projects Librarian established the workflow based on Vakil’s instructions for merging the spreadsheets using the Python script, a colleague mentioned an alternative method to accomplish the same merge using OpenRefine. The OpenRefine method involves pasting the contents of both spreadsheets into a common file, sorting the data by the common datapoint, and then using OpenRefine’s “Blank down” and “Join multi-valued cells” functions to merge the different sets of data based on the duplicate values. The Archival Projects Librarian successfully used the OpenRefine method on subsequent data cleanup projects.

Batch Updating Container Locations

The script that the USC Libraries used to change each container record’s location in ArchivesSpace only requires two data points: the container URI representing each relocated box and the corresponding location URI representing each destination shelf. The script that the USC Libraries used was adapted from a Python script that Mark Custer, Archivist and Metadata Coordinator at Yale University Library, shared in his initial response to USC’s call for help to the ArchivesSpace Users Group. The original Python script that Mark shared is titled “top-container-update.py.” The “top-container-update.py” script was written to make multiple edits to container records in ArchivesSpace, including location and barcode edits depending on each container’s existing data. Zahid Rafique, MIS Director for the USC Libraries Digital Library, adapted the “top-container-update.py” script to a simplified version that only edits each container record’s location. Zahid’s script, which he wrote using the Java programming language because of his familiarity with Java over Python, references a corresponding CSV file with just two columns: one titled “ContainerURI” and the other “LocationURI.” The last step in updating the locations of existing container records in ArchivesSpace was running the top-container-update script, which involves both connecting to the ArchivesSpace API and using the values in the CSV file as input.

One limitation of both versions of the “top-container-update.py” script is that the script does not retain each container record’s previous location in ArchivesSpace. After running the script, each updated top container will only show one location (the new location) in ArchivesSpace. Due to this limitation, the USC Libraries collected and saved previous container location data from ArchivesSpace before running the script, using the ArchivesSpace “Shelflist Report” and the data from the Get_Top_Containers Python script documented in the webinar presented by Vakil.

---

Saving these previous locations is essential for remediating mistakes made during the location update process and for tracking boxes that were not moved with the other boxes of their collections (e.g., boxes that were in use during the time of the move, but that had their locations updated because the movers reserved specific shelf space for them).

Figure 6. The three main sources of data combined for the top-container-update script.

**Batch Creating New Container Records**

Along with moving collections that already had container records in ArchivesSpace, the independent contractor also moved collections that had minimal collection-level records in ArchivesSpace without individual container records. In order to maintain (or in some cases initiate) physical control of these collections, the USC Libraries used a separate Python script and the ArchivesSpace API to batch-create new container records. As of ArchivesSpace version 3.1.2, there is no way to batch-create new container records with locations linked to the same resource or archival object record using the ArchivesSpace staff interface. ArchivesSpace users can use the “Load via Spreadsheet” tool in the staff interface to batch-create new container records, but not with linked locations. Container records created using the “Load via Spreadsheet” tool would need their locations configured following their import. Instead of using the staff interface, the USC Libraries used the ArchivesSpace API and a Python script developed by William David (Dave) Mayo, Senior Digital

---

Library Software Engineer at Harvard Library. The script is titled “import_container_data.py,” which “takes a spreadsheet with container data and ingests that data to the ASpace API.” The script can be used to create new top container records with various container details specified in the referenced spreadsheet, including container location, barcode, profile, and type.

The import_container_data.py script is designed to import new container records linked to specific archival object records referenced in the spreadsheet. However, since many of the collections that USC Libraries moved only had collection-level (resource) records in ArchivesSpace, USC used a slightly revised version of the import_container_data.py script—also developed by Dave Mayo. The revised version of the script could link the new container records to existing archival object records or resource records in ArchivesSpace. Linking the new container records to resource records worked well for relatively smaller collections (collections with fewer than fifty containers), but linking more than 50 container records to a single resource record slowed the loading of that resource in ArchivesSpace. Due to this loading issue, the USC Libraries started to distribute the new top container records among multiple “unprocessed” archival object records for larger collections. For example, a resource record describing a collection of 100 boxes would have two archival object records titled “Unprocessed boxes 1-50” and “Unprocessed boxes 51-100” with the new container records created via the import_container_data.py script divided among those two archival object records accordingly.

Challenges with Certain Forms of Archival Description

Given the relatively flexible nature of archival description, tracking containers and locations using pre-existing description in ArchivesSpace was not always straightforward. One challenge that came up frequently at USC Libraries was the formerly common practice of creating a single container record to describe a large range of boxes. Instead of having one container record per box, many of USC’s finding aids had one container record titled something like “Box 1-20” with a vague East Library location, such as “ELB, 1, cage” (indicating a fenced off area on the first floor of the East Library building). Past archivists employed this method as a workaround for minimally describing large parts of collections. Since these past archivists did not have an easy way to batch-create individual container records, they created these ranged container records as a shortcut. Working under the past limitations of information management systems like Archivists’ Toolkit and ArchivesSpace, these archivists prioritized the work of establishing intellectual control over the collection’s contents instead of ensuring the future physical control over the containers holding the material.

Following the move out of East Library, these ranged container records were no longer useful—especially for collections that were split up and shelved in different parts of the Grand Archives stacks. The work of replacing these ranged container records with individual records linked to specific shelving coordinates was relatively simple when a relocated collection only had a minimal collection-level “resource” record in ArchivesSpace. In such cases, the archivist could easily spot the ranged container records without having to explore multilevel hierarchies of archival object records. After finding a ranged “Box 1-20” container linked at the resource level, the archivist could use the import_container_data.py script to create 20 new individual containers—all linked to the resource record—and then delete the old “Box 1-20” record.

Unfortunately, remediating descriptive issues like the ranged container records became much more complicated and labor-intensive with more detailed finding aids. Some of the collections that moved from East Library to Grand Archives had detailed finding aids that went well beyond the collection-level resource record. Many of these collections had been processed at the folder- or box-level with corresponding “archival object” records categorized under complex hierarchies of series and sub-series in each collection’s finding aid. Finding and correcting ranged container records under these detailed and often inconsistent hierarchies of description was challenging for various reasons. The Get_Top_Containers script described earlier in this paper successfully retrieved all container records within a given repository regardless of where the container records were linked in each finding aid. But if a container record had to be replaced with new container records, the archivist had to link the replacement container records to the correct resource or archival object record(s), which could prove time-consuming for more complicated finding aids.

Using the import_container_data.py script to link new container records to a single resource record or a few broadly described archival object records did not require as much work because the archivist could reuse the same ArchivesSpace record ID for each new container. Conversely, in the case of a more detailed finding aid with smaller ranged container records (e.g., “Box 1-6”) scattered across various archival object records and nested under different sub-series, using the import_container_data.py script to replace these ranges with individual containers became more labor-intensive than the (also tedious) method of replacing the records manually in the ArchivesSpace staff interface. In order to preserve the correct links between existing archival description and corresponding container records, the archivist had to include each archival object record URI in the import_container_data.py spreadsheet to create and link the new container records to the appropriate place in each finding aid. In summary, many collections that had been meticulously processed and described were harder to document post-move than unprocessed or minimally Processed collections.
Post-Move Shelf Reading

After USC Libraries made as many container and location updates as possible using the independent contractor’s shelf lists in conjunction with the top-container-update.py and import_container_data.py scripts, the next step was to shelf-read the new archives storage in Grand Library. The Shelflist Report available in the ArchivesSpace staff interface was very helpful for this part of the process. The Shelflist Report includes a list of all ArchivesSpace locations and their associated top containers for a given repository. After exporting the report in a spreadsheet, USC Libraries isolated only the Grand Archives locations and reordered them by shelving range so that the person shelf-reading could walk the ranges with the report and note any discrepancies. The post-move shelf reading ended up being more important than anticipated after the Libraries discovered that the independent contractor who moved the boxes did not create shelf list spreadsheets for all of the collections that they moved. USC Libraries personnel had to gather locations for these collections during shelf reading.

Conclusion

The project to document the move of USC’s archival collections out of East Library was centered around an archivist’s perspective and skillset. The fact that this project was left to an archivist shaped many of the approaches and strategies used in documenting this move. For example, the decision to use ArchivesSpace as the main system of record for tracking the move was born (at least partially) out of an archivist’s familiarity with that system. In many respects, an archivist’s skillset was uniquely well-suited to the work of documenting this move. When one of the independent contractor’s shelving spreadsheets listed a box that was not previously documented in USC’s information management systems, the archivist could use ArchivesSpace to describe that box and link it to the appropriate place in an existing finding aid. In essence, the archivist had the knowledge and experience necessary to append new information arising from the move to the appropriate form of existing information (in this case, a finding aid in ArchivesSpace).

On the other hand, centering ArchivesSpace and existing archival description in the documentation of this move also slowed the project in certain areas. The issue of the ranged container records (past archivists describing many boxes with one container record in the finding aid) serves as a good example. Bundling the tasks of remediating existing container data in ArchivesSpace with the project to document the move ended up slowing both of these projects. Furthermore, the timeline for documenting the move out of East Library was a concern for many stakeholders, especially because the capacity for multiple Libraries units to fulfill researchers’ requests depended on this documentation. Using the ArchivesSpace API to batch-edit existing archival description highlighted some of the disconnects between the labor involved in establishing intellectual control (describing the informational content of archival resources) versus maintaining physical control (tracking physical locations of
archival holdings). Archivists need to be aware of these potential disconnects starting in the early stages of accessioning and describing archival resources. Enhanced training and procedures manuals outlining the consequences that certain descriptive practices have on the future maintenance of physical control could help address some of these disconnects.

According to Charlotte Priddle and Laura McCann’s 2015 study on offsite storage and special collections libraries, the use of offsite storage by research libraries has increased significantly since the early 1990s. A 2006 Association of Research Libraries (ARL) survey cited in Priddle and McCann’s study revealed that 51 (65 percent) of the 79 ARL libraries that responded to the survey store special collections holdings in access-restricted offsite storage facilities. This increased use of offsite storage, paired with the fact that archivists often act as the main administrators of archival management systems like ArchivesSpace within a given repository, adds to a growing set of responsibilities and required skillsets for maintaining physical control over archival holdings. The tools and practices that archivists use to maintain physical control need to adapt to these changes. Ideally, archivists would be able to batch-edit and batch-create container records and shelving locations more efficiently in systems such as ArchivesSpace without having to train themselves on accessing APIs and writing Python scripts. Until archival management systems can deliver these enhancements, archivists will need to continue to juggle various roles, including accessioning and processing archival material, accessing APIs and running Python scripts to update archival data, and advocating for system improvements among user and technical advisory groups.