

8-14-2019

Out With the Old, in With the ... ArchivesSpace

Laura Buchholz
Reed College

Recommended Citation

Buchholz, L. (2019). Out With the Old, in With the ... ArchivesSpace. *OLA Quarterly*, 25(1), 30-33.
<https://doi.org/10.7710/1093-7374.1973>

© 2019 by the author(s).

OLA Quarterly is an official publication of the Oregon Library Association | ISSN 1093-7374

Out With the Old, in With the ... ArchivesSpace

by **Laura Buchholz**
Discovery & Systems Librarian,
Reed College
buchholl@reed.edu



LAURA BUCHHOLZ is the Discovery and Systems Librarian at Reed College, where she previously served as the Archivist and Digital Infrastructure Librarian in Special Collections and Archives. Her research interests include user interfaces for archival materials, accessibility in library systems, and highlighting invisible labor in libraries and archives.

What do you do when your cataloging infrastructure is outdated, unusable by your patrons, and creating excess work for your staff? Get new infrastructure! At Special Collections and Archives in the Reed College Library, we are in the process of implementing ArchivesSpace, a collection management tool for archives. As we go, we are professionalizing our description practices, preparing to contribute to regional aggregations (Archives West), and, most importantly, increasing access to our collections. The software is also easier to use and reduces the time we spend doing research for our patrons.

Archival materials can be challenging to catalog, and confusing for patrons to navigate. Where a book will have a clear title, author, and date of publication, materials found in archives are the byproduct of daily life and business. They are usually created without any intention that they will end up in a library catalog or an archival database. Archival materials are also context-dependant, meaning that a single page or document might make no sense on its own. Knowing a little about the surrounding materials can help, and knowing a little about the creator might also help. Imagine a photograph of a group of people. You might not know who they are or what they're doing. But if the next photograph shows the same people, in the same clothes, cutting the ribbon for a new building, then you can infer that the first photo was also of the groundbreaking of that building. For this reason, archival materials are usually described in finding aids, rather than in a catalog like books.

Our existing infrastructure consisted of finding aids in PDF format, linked from our web page by last name of the collection's creator. Patrons could view them, but couldn't easily search across collections. The other major component was a FileMaker database, set up in 1991. Now, there's nothing wrong with FileMaker as a piece of software! But this database was only set up with records for individual folders of archival materials, or sometimes for individual items. There was no collection-level information. Imagine a library catalog without records for any books, but with thousands of records for book chapters and sometimes even pages.



To find the physical folder or item from a record in FileMaker, staff had to memorize where to go. Sometimes, we would find something in the database that was interesting or relevant but not know where to find it. The FileMaker database was set up before any professional archivists worked in Special Collections and Archives, and we struggled to implement standard descriptive practices in the database.

Even more problematic was that the FileMaker database was internal-use only. When patrons visited, they might describe their question or what they hoped to find. Special Collections and Archives staff would then search the database and pull out what we thought was relevant. Patrons couldn't browse, couldn't search, and couldn't evaluate for themselves what was relevant and what wasn't. As librarians and archivists, this didn't sit well with us. We don't collect archival materials only for the sake of having and preserving them; we collect them to be able to provide access! In addition to ideological concerns, we simply couldn't keep up. For a small school, we were relatively well staffed at 2.5 FTE, but we struggled to do research for patrons as well as complete our other duties.

Our goals in leaving our PDF + FileMaker infrastructure were to (a) allow patrons to search and browse descriptions of our holdings, with our assistance only when they needed it; (b) to reduce the extremely high institutional knowledge needed to search and locate our holdings; and (c) to describe archival materials following best practices.

Selecting a system was easy: ArchivesSpace is a standard tool used in many archives in the United States. More importantly, it is supported by our regional consortium, the Orbis Cascade Alliance, so we knew we would have workshops, training guides, and the support of many colleagues. ArchivesSpace is free and open source, and we are lucky to work at an institution with an IT department friendly to open source software. We had a locally installed and maintained instance of ArchivesSpace without much trouble.

To migrate, we started with the easy stuff: the 70 or so finding aids available in PDF format. Although these did not follow current archival standards—using EAD and DACS—they were close enough that there wasn't much trouble. Initially, I tried to automate the ingest: saving the text of the PDFs to various plain text and CSV formats, then attempting to systematically clean up the data into a format that could be uploaded to ArchivesSpace. The PDFs had been created in various versions of Microsoft Word, though, without a standard template or formatting. After a week or two of messing around, I asked one of our student employees to simply copy and paste the data of a finding aid into a new record in ArchivesSpace. Five minutes later, task complete, they asked what else they should work on! I was hoping for a less manual process, but realized that sometimes the simple option is the best option. Two student employees, each working about 6–8 hours per week, had the PDF finding aids in ArchivesSpace within two weeks.

The FileMaker database has proven to be more challenging. Although it was easy to export data, the data itself frequently lacked context. ArchivesSpace is made to support current archival description standards, the most important of which is to have information about the collection: what it contains, how much stuff it contains, how it is organized, and who created it. That is exactly what we didn't have. Instead of leaving those uncertainties in place, we started to create these collection-level descriptions so that we would have a place to put the records from the FileMaker database. Where information was lacking, we re-



corded what we did know and were transparent about what we didn't. Midway through this process, one of our staff received training on DACS (the archival description standard), so that they were able to write these descriptions according to best practice. After some training, our student employees were able to help as well.

As collection descriptions were being established, I started working on attaching folder and item descriptions exported from FileMaker. The export itself was simple, as the database was only one table, and search sets could be easily saved to Excel. With several rounds of iterative searching, I was able to get sets of records that mostly belonged to the new collection descriptions.

In the exports, it was fairly easy to see what fields and values were no longer needed—for example, fields with no values or information that had been out of date for years—but there was one big problem: almost all information about a folder or item had been entered into the Title field. A typical title might contain the folder or item title, subtitles, alternative spellings, dates, keywords, relevant names, notes about where the information came from, and sometimes even transcriptions of the content of the folder. All this in one title field and with inconsistent formatting! This had been done to ease cross-field searching (by eliminating the use of other fields) in FileMaker, but made it nearly impossible to easily migrate data from FileMaker to ArchivesSpace.

We also still faced the problem of finding a record in FileMaker, but not really knowing where to find a hard copy—or if it even existed physically. Because one of our goals was to reduce the high levels of institutional knowledge required to find resources, we decided to take on the additional work to fix these problems, instead of simply migrating the existing data. After export, I would attempt to sort the records to match what the arrangement was: usually alphabetically, sometimes chronologically. Then, I would review the export against the physical boxes, noting what boxes the folder record was in, or if it was missing, and fixing the order of the records as needed. This was, and continues to be, the most time-consuming part of the process. Imagine if your library catalog contained titles of books, but no call numbers, and the books were organized on the shelf by call number. You would either need experts to translate titles to call numbers, or would need to go through the painstaking process of connecting titles and call numbers in the catalog.

After the export is sorted and box information is recorded, the remaining steps are to label the physical boxes and upload the data to ArchivesSpace using Harvard's "aspace-import-excel" plugin. Finally, I mark the records as "migrated" in a newly created field in the FileMaker database. At the end, I expect to have a few hundred random records left in FileMaker that do not easily fit into a collection, which we plan to review later on a case-by-case basis.

After nearly a year and a half, we've migrated about 90 percent of the collections to ArchivesSpace, with one archivist working an average of about 70 percent of time on the project, another at about 10 percent time, and with an archives specialist dedicating about 30 percent. Our student employees, as always, are crucial to our success, and have contributed about 40 percent of their total hours. Most of the time spent has gone to fixing discrepancies between the FileMaker database and the materials in the pre-migration stage.

We turned on the public interface for ArchivesSpace early, after only a few months, which allowed people to search what we have migrated. We continue to point our patrons



there and have already noticed an improvement. Almost every spring, a history class has an assignment where they need to research a topic in the college archives. It has always been a fun class to help with, but it can take a lot of our time to search for relevant materials for 10–20 students, all on the same deadline. It has also been frustrating, because students get a skewed version of archival research in which the archivists and staff translate their queries into boxes and folders magically pulled from the stacks. Last spring, we started showing students how to search our resources using ArchivesSpace, explaining the components of a finding aid as we went along. Students pick up quickly, and seem to appreciate the ability to do part of their research outside of our open hours.

Although we still have more work to do before the migration is complete, we have learned a few lessons along the way. Finding a good balance between fixing old errors and getting the project done can be challenging. The first collections that I tackled had been processed in such a way that, to migrate them, I essentially ended up reprocessing them. That quickly sucked up available time, and I learned to only fix what would be difficult and very time-consuming to fix post-migration. Another lesson was a reminder to be flexible. Since the FileMaker data is so easy to export, I had originally thought it would be the simplest part. The lack of consistent structure to the data in the title field, along with the inability to identify physical locations without a manual check, made this stage take much longer than I had planned. Finally, ArchivesSpace is not a perfect piece of software. We knew that going in, but the public interface leaves much to be desired from a user experience perspective. For example, in finding aids, each level in a hierarchy inherits information from the levels above it, allowing archivists to avoid repeating information. In practice, patrons are sometimes confused by what they see in component-level records, especially when they arrived at a component via search instead of browsing the finding aid from the top. Still, a sub-optimal user interface is better than no user interface at all.

As we move to the final stages of the migration, an increasing number of our collections are browseable and searchable by patrons. We still provide assistance and consultation, but now are able to let patrons do the bulk of their research for themselves. Our student employees are able to find materials without relying solely on staff, and we hope to have reference and instruction librarians take regular turns at our reference desk. Lowering the level of institutional knowledge required to provide reference assistance in Special Collections and Archives helps greatly as personnel move to different positions or leave the library for other opportunities.

Most migrations are a challenge, and moving from PDFs and an outdated FileMaker database to ArchivesSpace has been no exception. The good news is that the benefits are already noticeable, and more are on the horizon. 🐼

