

ArchivesSpace™ Adventures: A Migration

Adam Smith, Archives Technician

and

Heidi Southworth, Digital Initiatives Librarian

Minnesota State University, Mankato

University Archives and Southern Minnesota Historical Center

PO Box 8419

Mankato, MN 56002-8419

507-389-1029

507-389-5066

adam.smith-3@mnsu.edu

heidi.southworth@mnsu.edu

An Archival Journey

On February 4, 2019, the University Archives and Southern Minnesota Historical Center at Minnesota State University, Mankato successfully launched a new archival tool for our patrons called ArchivesSpace. While this at first glance may not seem like a big deal, the journey that the archives undertook to transform this search tool from a set of static HTML pages (all 700+ of them) to an easy-to-use search engine contained danger around every corner. The team had to fend off lions, tigers, and bears and had to blaze a path through a thick forest of metadata and archival records. The journey traveled down a dark and scary path; the path less traveled. At the path's end, the archives team used their superpowers and a little magic, and thus emerged triumphant with a bright and shiny new archival tool called ArchivesSpace. Okay, so maybe this is exaggerating a little bit. There were no lions, tigers or bears, no forests, no superpowers, no magic, but like any good story, this one has a great ending. A small archives unit from a mid-sized university archive used teamwork, investigative know-how, learned from others, and partnered with those, who had some serious technology skills, to transform all those webpages into a better search experience for the patrons. The journey that follows speaks of how in collaboration with the library systems team, archives staff successfully built a local ArchivesSpace instance.

Now, do not quickly judge this article as being against the company Lyris; this is far from the truth. Lyris is an academic consortium that specializes in open technologies and community supported software for libraries, archives, museums, and similar organizations. The archives team is indebted to Lyris and others in the community for developing ArchivesSpace, but the archives just simply could not take on a membership fee at the time. The team worked with no

budget and relied on individual team members to see the project to completion. Someday the situation (budget or otherwise) may change, and the archives will readjust as needed. Individual situations may be different, so the archives team encourages others to talk to Lyrasis as they may have options for a particular archives or institution. It is the hope of the authors that the following information will at least provide some helpful information on how a small team built a local instance of ArchivesSpace and how that team is using it to create a better, more comprehensive search tool for archival materials.

The University Archives and Southern Minnesota Historical Center at Minnesota State University, Mankato

Minnesota State University, Mankato is situated in south central Minnesota about 85 miles southwest of the Minneapolis/Saint Paul metropolitan area. The University, founded in 1868, serves approximately 13,000 to 14,000 students with 1600 faculty and staff. Minnesota State Mankato is a regional comprehensive university and in addition to bachelors' degrees, grants masters and doctoral degrees as well. Memorial Library, built in 1967, houses the University Archives, which serves as the official repository of the records of the University and was informally established in 1961-1962. In 1979, the opportunity arose to create a more formalized and organized University Archives, thanks in part to the dedication and efforts of history professor, Dr. William Lass, who is to be credited with saving many of the University's records from as far back as 1868.ⁱ The University Archives also serves as the Southern Minnesota Historical Center (SMHC). Established in 1969, the SMHC is one of several Minnesota regional research repositories and contains manuscript collections from twelve counties in south central

Minnesota: Blue Earth, Brown, Faribault, Freeborn, Le Sueur, Martin, Nicollet, Rice, Sibley, Steele, Waseca, and Watonwan counties. The Center contains organizational and business records as well as personal papers dating to the 1860s. For many years, librarians and library staff cared for the University Archives, but there was no archivist until 1999 when Library Services hired a full-time certified archivist. In the beginning, Archives staff remained at two positions, the archivist and an archives technician, before growing to include a digital initiatives librarian in 2013 and a second archives technician in 2014.

In those formative and subsequent years, Archives staff kept track of accessions and collections in a variety of ways. With much to do, they recorded early accessions on a simple paper notepad. As archival collections were created, staff developed a Microsoft Access database to keep track of these materials. Over time, the archives team created archival finding aids or collection guides following archival standards. The public could access these finding aids on over 700 HTML coded webpages linked through the web site. In addition, the cataloging and metadata librarian developed MARC records for the library's catalog that allowed the public and now a global audience, a glimpse into what archival collections contained. For staff in the archives, however, more detailed information about the collections resided in those Access databases and simple notepads.

The University Archives' first extensive database of archival information came in 2012, as staff developed a local instance of ARCHON. Hosted by the library's system team, the ARCHON

database contained accessions and basic collection level information, such as box locations. More detailed information, such as folder level information, remained on those 700+ webpages. While ARCHON gave archives staff a new method of organizing collections and accessions, the archives did not use ARCHON to provide the public with any information. A truly behind-the-scenes product, patrons had to view finding aids for folder level information and did not use ARCHON. Archives staff maintained the information in multiple locations: on webpages, MARC records, and ARCHON database, and if anything changed, the changes needed to be made in all the separate locations. While a Google search box on the main archives page helped to search across all finding aids and webpages, the archives needed one tool that would help maintain the collection and provide the public with a more comprehensive search engine.

One More Migration?

Starting in the 2017-2018 academic year, the archives and the library faced multiple technological changes on the horizon. First, the systems team within the library managed the library's servers, including the server that hosted ARCHON. In 2017, the University's Information Technology Solutions (ITS) asked the library systems team to physically move the library's servers to the ITS Computer Center for security purposes. In addition, ITS informed the library's systems team that the software utilized for the servers, VMware, needed to be decommissioned and that the servers be moved to a Microsoft Azure or Windows server. While the silver lining in all this meant that the library systems team would receive a server and disk space, it also meant that ARCHON would need to be potentially moved to continue its use.

Moreover, since ARCHON had limited vendor support, ITS considered it to be a security risk and encouraged an examination into its future use.

Secondly, in 2017 and 2018, the library and archives faced an enormous task with the launch of the University's new CMS or Content Management System for all university-based web pages. Responsibility for updating all archives pages fell to the archives team, which was temporarily understaffed. The archives web pages not only included several standard informational pages but also all those 700+ finding aid pages with folder level information formatted into nice, neat tables. The new CMS, however, did not work so well with tables, and all 700+ finding aids would need to be manually formatted into something other than a table. This task of migrating these finding aids, however, would exceed the time allotted for the migration. Therefore, as archives staff began to plan the work on the remaining archives pages, which would go live in August 2018, converting 700+ finding aids did not appear to be a good option.

At the same time archives staff were learning a new CMS, the library began planning for a migration to an entirely new integrated library system. The library operated on ExLibris Aleph, and the decision was made to migrate to ExLibris Alma/Primo in January 2019. The library's migration team formed in 2017 to not only review the options for a new system but to also learn it, figure out all the customization, implement, and then fix everything that did not go as planned. The archives team was involved in this process since the beginning, and so had additional workload to complete. Archives staff offered input, tested the new system, and double checked

the collections and metadata that the archives oversaw. The archives staff had a long and growing list of projects, but instead of decreasing the list, staff decided to add to it. The staff in the archives kept thinking about how the public could not really search the collections through ARCHON. What if the archives had a system that could be used to track accessions and inventory and offer collection and folder level information for the public all in one interface? So, the archives decided that ARCHON needed to be migrated to a new user-friendly platform.

What? Yes, in the middle of two major migrations, archives staff decided to add another. One would probably question the archives' planning skills at this point and say that the timing did not quite add up. Archives staff, on the other hand, thought, why not? In hindsight, archives staff would recommend that maybe one should think carefully before taking on three migrations at once, but the archives and all its information survived despite the decision to conduct multiple migrations at the same time. Archives staff started exploring alternative options right away and utilized knowledge from previous experiences to create a list of available products. The Society of American Archivists has an excellent page of archival technologies and resources that staff reviewed as well.ⁱⁱ Archives staff first examined proprietary based software. While many of the products would suit the needs of the archives, many of the membership costs or associated fees that came with some of the products were beyond the archives' budget, which basically stood at zero. The archives, however, did have an excellent technical support structure within the library that could help with technological details and programming. As a result, archives staff worked with systems staff to look for options that required more technical knowledge but came at zero cost, such as an open-source product.

There are many open-source products, but archives staff did not have much time to devote to examining all of them. Instead, the team focused on a more detailed examination of just a few products, namely Collective Access and ArchivesSpace. Before coming to Minnesota State Mankato, one archives technician worked in the local history museum field and utilized Collective Access. Collective Access provides robust collections management solutions for museums. Archives staff examined Collective Access for use as there were 3D artifacts and art collections as well as traditional print materials in the archive's collections. Because of the size and scope our collections, however, archives staff decide that ultimately Collective Access just did not fit what the archives needed. Archives staff then focused on the one that was ultimately selected-ArchivesSpace by Lyris. An open-source product, ArchivesSpace is used by several, larger archives and would be able to handle the archives' collections. In addition, the product would provide a robust, easy-to-use, public search engine for archives' patrons. Unfortunately, archives staff lacked the technical expertise to install the software, and at the time, the archives could not afford a membership through Lyris to access any documentation or manuals. What happened next? Since the option to run around screaming for help would disrupt the archives and library patrons, archives staff instead turned to an internal support team.

Technical Jargon and Magic Section (Proceed with Caution)

The archives received support from a team of smart information technology specialists (IT specialists) led by a systems librarian. The systems team made the migration to ArchivesSpace

possible. They took all the technical jargon and turned it into a wonderful new magical search engine. The ArchivesSpace project originally started out as an internship project for one of the IT specialists. The IT specialist, who helped develop the local instance of ArchivesSpace, remarked that while he had a little knowledge about web servers and Linux, he learned much as the project progressed, and he started the project by first looking at the requirements for ArchivesSpace.

Information on these requirements for ArchivesSpace can be found in GitHub.ⁱⁱⁱ In short, the basic ArchivesSpace requirements (at the time of the migration) were that one uses a supported operating system such as Windows, LINUX, or Mac OS X and Java 1.8. A system should have at least 1024 MB ram allocated to the application with at least 2 GB for optimal performance. MySQL was not required but was strongly recommended for production use. Now, how many people are now lost by all this technical jargon? Anyone? If you are lost, do not worry. Archives staff were a little lost, too. After all, archives staff are specialists in the archives. While staff understand technological things like digitization, digital collections, and digital preservation, the archives team is still a bit mystified when one starts talking about LINUX, web servers, and Java, which does not refer to delicious coffee in this case. If you feel like this, do not worry. You may have partners like IT support on your campus or in your organization that you can turn to for help and you can always talk to LyraSIS for their technical expertise as well.

The IT specialist next had to update the archives' ARCHON server from 3.21 rev 1 version (the last official release) to a community release 3.21 rev 3 version. This update would allow the IT

specialist to use the migration tool developed by ArchivesSpace to move our data from ARCHON to ArchivesSpace without too much hassle. The IT specialist ran many tests before the final move to make sure that there was no data lost during the actual move. Again, it is to his credit that the IT specialist put up with archives staff asking a million questions and hovering anxiously in the background as he worked with the data. In the end, archives staff did not need to worry. The system team who handled all the technological components made the migration so smooth and successful - just like magic.

The key to our success was testing, lots and lots of testing before launching our final product. The IT specialist used Windows WAMPP (Windows, Apache, MySQL/MariaDB, PHP, Pearl) to test ArchivesSpace. Ultimately, he decided to utilize with Debian LINUX because it was light weight and supported by the campus IT Solutions department who were going to host it. While the archives and library have their own system team, the systems team worked closely with the campus IT Solutions department. IT Solutions requested Debian LINUX because they would be maintaining the server, conducting the updates, and providing offsite and onsite backups and security. The library systems team would just be responsible for managing the ArchiveSpace application and configuration.

In the end, here are the products that the IT specialist used to create our instance of ArchivesSpace. He used Debian LINUX, MariaDB 10.1.44, Java 1.8.0_232 and Apache 2.4.25. He also set up a test instance using a Windows 10 machine running a virtual machine using

VirtualBox. This way, the specialist could continue to test and make changes in the virtual machine before coordinating with IT Solutions to update and change the production server. Now, if you have gotten this far without being completely confused, congratulations! Archives staff understood a little bit about the process and could mostly follow along, but the team never had to worry as there was lots of support along the way. The systems team has continued to maintain and update our ArchivesSpace instance, and they have already gone through a design change to the home page. The work was not completed when the ArchivesSpace instance went live, however, as there were still many projects to do.

Post Migration

During migration, archives staff realized the need to complete several projects to make ArchivesSpace functional both for employees on the back end and for a public interface on the front. Unfortunately, because of the University website migration and losing access to the finding aids, patrons needed immediate access to that information. The first post migration project was ensuring public access to basic finding aid information. It took two steps to accomplish this task. First, student workers digitized and created OCR enhanced PDF scans of the paper finding aids. Student workers then uploaded the documents to ARCH, the University Archives digital collections repository. This step ensured patrons and staff had full-text searchable finding aids rather quickly in the project.

The second step in ensuring public access to the collection also involved student workers. The student workers copied and pasted basic finding aid information into ArchivesSpace such as titles, abstracts, historical/biographical information, and scope. The students' work enabled patrons to discover the collections in ArchivesSpace through traditional search engines such as Google. In fact, researchers all over the United States, such as South Dakota and California, and across the globe (Wales) are now able to find the archival collections and send direct questions to the team on what they found using ArchivesSpace. The prospect of reaching additional patrons through the inclusion of additional metadata is exciting!

Adding metadata is just one part of developing a good search engine, however, for one also needs to ensure that the information is correct and consistent. Due to the timing of the migrations, the archives team did not have a chance to do metadata cleanup in ARCHON before the migration. As a result, there are multiple clean-up projects to work on in ArchivesSpace in addition to adding the information found only in the finding aids. Some of these projects are basic in nature, such as doing an inventory, cleaning up location information and standardizing metadata, while others are more complicated and time consuming, such as adding box and folder level information and creating subjects and agents.

The first cleanup project the archives team developed established physical control of the collection. Student workers conducted a physical inventory of both the processed and unprocessed parts of the collection, and then compared the results to the location information in

ArchivesSpace. Archives staff rectified anomalies, such as missing boxes and misplaced items, and at the same time standardized the location metadata. Standardization involved adding leading zeroes to the metadata for physical location profiles, so they all had two or more digits depending on the collection's size. Staff also deleted duplicate and temporary locations no longer in use. The inventory project allowed the archives to reset and start over with a clean slate that is (mostly) fresh and cleaned up.

A second project to establish intellectual control of the collection and to aid patron research required archives staff to create subject and agent (such as peoples' names, family names or corporate entities names) records in ArchivesSpace. Staff used Library of Congress Subject and Name Authority files and included local sources from previously cataloged library records to create the subjects and agents. Staff then linked the subject and agent records to the Archival collections. Creating these links allows patrons to do a general search and browse the collections by using agents and subjects. Patrons can also discover links to other collections when looking at individual agents and subjects in the records. There is a considerable amount of work left to do both in adding subjects and agents, but also in editing the migrated files. The migration created hundreds of agents that are accession numbers rather than actual agents. Staff will need to delete those to make the real agents discoverable.

A third project archives staff (along with a student worker in the summer) are working on entails adding series, box, and folder level information from the finding aids. When editing in

ArchivesSpace, series and folders appear as children, grandchildren, and siblings, depending on what level one is at in the database. This project is only in the beginning stages and is expected to take a considerable amount of time. When completed, having this level of detail will allow staff and patrons full access to our collections through ArchivesSpace. Until then, staff will continue to direct patrons through the external link feature in ArchivesSpace to look at the finding aids in ARCH: University Archives Digital Collections that contain the series, box and folder level information.

Future Potential Projects

Archives staff are excited about the possibilities of future projects in ArchivesSpace. It will take a considerable amount of time to continue adding box and folder level information for the existing collections.^{iv} There is also the reality of adding metadata including extensive folder level information for future collections that need processing. While looking at these projects, Archives staff is exploring how to integrate future projects into our work plans. It is important to find a balance in tasks of improving ArchivesSpace along with processing collections, assisting researchers, and working with students. One option is doing small scale tests and then scaling them up with staff and student worker time to complete the projects.

A project that archives staff are particularly interested in is barcoding all of the archives boxes. Barcoding provides access to usage statistics so staff can efficiently track how patrons are using the collections and what items to prioritize for future digitization projects. Barcoding might also

make it easier when adding instances (locations) to the objects by being able to scan the barcode on the box into the ArchivesSpace field instead of manually browsing the collection to find it. A potential downfall is oftentimes boxes hold multiple collections and as a result the barcode would apply to multiple collections. Archives staff will need to explore this further to analyze the benefits and negatives with barcoding.

A supplemental project to barcoding the boxes is creating container profiles. Container profiles provide detailed dimensions of a storage container, while the “container types” currently used provide only a brief description. Utilizing container profiles will also allow the archives to track how much space is available on shelves based on those container profiles. Having that information allows the archives to optimize storage and have a knowledge of the best place to store or look for storage locations for boxes. There is also an option for creating location profiles. Archives staff have not reviewed this option or explored how it would work, but the expectation is it would operate similarly to container profiles and help optimize storage for archival collections.

The Unexpected but Beneficial

With any migration or project, there is always the unexpected. While the migration went very well, there were a few things that archives staff needed to address, and the migration has led to some great conversations and thought-provoking discussions. For one, terms were different than what archives staff knew from ARCHON, and so there was a bit of a learning curve. Terms such

as instance, agent, child, grandchild, and sibling had archives staff a bit perplexed at first. One thing that staff did early on was to simply do some web searching for other ArchivesSpace instances to see if they had any tips or best practices. While there are several, archives staff found Yale University's documentation extremely helpful to our projects.^v In case you are interested, for archives staff, an instance is a box; an agent is a person's name, and child/grandchild/sibling are all structural terms in the organizational hierarchy that is ArchivesSpace.

Additionally, staff utilized *Describing Archives: A Content Standard* (DACS) and other glossaries to help interpret terminology in ArchivesSpace. DACS helped archives staff figure out the terms used in ArchivesSpace, such as bulk date vs inclusive date (see DACS 2.4).^{vi} One of the helpful tips in ArchivesSpace is that if you hover over a field such as data type or the extents portion, an informational box will appear to describe DACS standards and how to use that field.

Archives staff debated questions that had not been anticipated at the start of the migration. For example, what do you call a box? (No, this is not a joke.) Standardized terminology is important in the development of a resource like ArchivesSpace. If a person entering the data calls a box an archives box but another person calls it a records box, the metadata will be inconsistent and potentially confusing to staff. Early in the project, archives staff standardize the terminology and discussed how to enter the data. For terminology, staff utilized the SAA Glossary, and if the term was not available through SAA, staff agreed on what to use.^{vii}

Next, what happens when you take a former cataloger turned digital initiatives librarian and put them in the archives? (Again, not a joke.) This former cataloger and others liked to talk about MARC subject and authority records and standardization of names and subjects. A second learning curve came in the form of archives staff using the Library of Congress' Authority database to generate uniform headings for names and subjects in ArchivesSpace. With ARCHON, archives staff had not worked with name or subjects in our ARCHON records. Now, staff had the opportunity to formalize names or subject and make them searchable and browsable in ArchivesSpace. One great feature of ArchivesSpace is the ability to click on Subjects or Name and have an actual list for patrons to browse. If these terms are all uniform and linked together, the end user will have a better and easier search experience.

In the end, all these discussions on terms and standardization had archives staff thinking about the entire processes. How were call numbers created? Should one add leading zeros to numbers in collections? What names do the collections have? Are they consistent and clear? While archives staff have not finalized everything, this migration allowed staff to think and discuss the small details that make up archival collections. While this is nothing earth shaking, it has helped archives staff to examine better processes, to think about consistency, and will lead to better documentation in the future and a better experience for the patrons.

Looking Ahead

With any migration, the work never seems to be complete. Data entry is an ongoing process; the systems team is doing on-going maintenance and updates to the site; and archives staff are also looking ahead. Archives staff have not done Encoded Archival Description (EAD) for the finding aids. Should this be done? While the answer is uncertain currently, staff did find out that one can export XML records in ArchivesSpace for entry into a system like WorldCat.

Additionally, archives staff recently completed a project where they checked to make sure that there are MARC collection level records for everything in our local catalog and in WorldCat. Staff found that several collections did not have MARC records created, so the library's cataloging and metadata librarian is working to add these records to the catalog based on the XML records that were generated from ArchivesSpace.

Another interesting and someday project is to see what ArchivesSpace can do for archival digital collections. Currently, the link in the archives' ArchivesSpace just directs patrons to the home page of the archives' Islandora repository or ARCH. A patron must then search ARCH for information, but it would be nice to have all the metadata for our digital collections searchable in one place. Again, archives staff do not have the answer, but as this would benefit the patron, staff are very curious to continue to explore this process.

In addition, work in ArchivesSpace has enabled archives staff to think about how we can handle special indexes to some of our collections. In particular, the Thomas Hughes collection is one of those special archival collections that contains 69 notebooks with research notes.^{viii} Hughes, a

well-known historian from Blue Earth County in southern Minnesota, kept these notebooks on his research, and added his own reminiscences on Mankato and the various people he met. These notebooks contain a wealth of information for genealogists and researchers. To help patrons looking for specific information in these notebooks, an index was created and added to the finding aid. As staff worked on adding this finding aid to ArchivesSpace, however, staff were not sure how to handle the index. Through trial and error, one staff member found a way to include the index with each notebook by adding an Index note. Now, the index is also searchable in ArchivesSpace. Archives staff have other indexes that can now be added to ArchivesSpace in a similar fashion and because of the versatility and capabilities of ArchivesSpace, staff cannot wait to figure out what can be done next.

The migration to ArchivesSpace was a lengthy process with lots of team members pitching in to build a resource with all the contents in one place. Well, staff may not have everything entered yet, but eventually, everything will be in that one central place. This centralized search engine has made all the archives' physical collections easily searchable. Patrons can search before they even visit the archives. In fact, one patron recently sent archives staff a list of exactly what boxes they wanted to view ahead of time, and staff were able to have everything ready before the researcher even stepped into the archives. Patrons have also discovered our new "request" feature in ArchivesSpace that sends archives staff an email asking about a particular collection.

While all this is making life easier for patrons to see what the archives has, it has also made it easier on the archives' staff. Archives staff are planning on utilizing ArchivesSpace for all aspects of archival work from accessioning to finding aids. For example, in using ArchivesSpace to create accession records for one such new collection, the COVID-19 Community History Project, archives staff have greatly benefited from the use of ArchivesSpace to help organize and make sense of the hundreds of donor forms and submissions that have a variety of restrictions. This large collection's accession records are now very well structured and organized and restrictions are quickly identifiable and visible to archives staff. This organization has helped staff to quickly find and identify not only what is on the shelf but what requires special handling.

Archives staff are continuing to work on completing box and folder level information in ArchivesSpace. When finished, archives staff will have one place to search physical collections rather than multiple pages. Staff will have a better understanding of the exact contents of all the collections and will be able to quickly respond to requests from patrons. The ability to now search by subjects and by people's names will make it even easier to find a specific thing in a collection and in a relatively short amount of time. No more searching here or there or on 700 webpages. In the end, the benefits of migrating to ArchivesSpace are making work a little easier and a lot more organized. While the COVID-19 pandemic has made the archives a challenging place, archives staff are confident that the new ArchivesSpace tool will lead to even better days ahead.

The Minnesota State University, Mankato University Archives' ArchivesSpace instance^{ix} is available for browsing and searching and staff would welcome any comments or suggestions.

Please contact archives staff with questions or comments.^x

ⁱ Lass, William E., *Minnesota State University, Mankato, 1868-2018: A Sesquicentennial History* Minnesota State University, Mankato, 2018, 190-191.

ⁱⁱ See <https://www2.archivists.org/american-archivist-reviews/tech-and-resources>

ⁱⁱⁱ GitHub, <https://archivesspace.github.io/archivesspace/user/getting-started-with-archivesspace/>

^{iv} Lisa Calahan & Kate Dietrick (2016) Setting the Stage and Keeping Sane: Implementing ArchivesSpace at the University of Minnesota, *Journal of Archival Organization*, 13:3-4, 114-126.

^v Yale University Archives, <https://guides.library.yale.edu/archivesspace/home>

^{vi} Describing Archives: A Content Standard, <https://www2.archivists.org/standards/DACS>

^{vii} <https://dictionary.archivists.org/>

^{viii} Hughes, Thomas, 1854-1934, and Family. Papers, 1855-1946. SMCH Manuscript Collection 101, Southern Minnesota Historical Center, Memorial Library, Minnesota State University, Mankato.

<https://arch.lib.mnsu.edu/islandora/object/MSUrepository%3A155176/datastream/OBJ/view>

^{ix} Minnesota State University, Mankato University Archives ArchivesSpace instance, <https://archivesspace.lib.mnsu.edu/public/>

^x Minnesota State University, Mankato University Archives contact information, <https://library.mnsu.edu/archives/about/contact/>