

**Final Report**  
**ACMS Pilot 2015**  
**Briscoe Center Archival Description Task Force**

**Executive Summary**

In light of recent efforts on campus to centralize planning, infrastructure and services across the university, members of the special collections community are currently looking at ways they might combine efforts towards ensuring secure and preservation-minded physical storage for analog primary resource materials. At the same time, special collections across campus are also actively engaged in digital infrastructure building, resulting in conversations about how best to create, manage, store, update and publish our collection data so that we are 1) increasing discoverability of our collections and the efficacy of our life-cycle management while reducing the average time spent managing our collection data and 2) leveraging the power of our collection data to inform decisions about processing, conservation and collection development.

In the spirit of these broad-reaching changes, the Briscoe Center sponsored a pilot project in Summer 2015 to perform a high-level analysis of our collection data and to test ArchivesSpace as a possible archival collection management solution for the Center. The pilot work was completed by the Briscoe Center Archival Description Task Force, a group including representatives from Archives & Manuscripts, Photographs Archives and Digital Archives. The Archival Description Task Force (ADTF) charter is as follows:

- Review state records retention and preservation requirements for collection data kept and maintained by campus special collection repositories
- Document current archival descriptive workflows and associated pain points
- Investigate methods for streamlining archival descriptive workflows
- Conduct peer research to estimate resource commitment necessary to implement an archival collection management system
- Identify implications of changes to archival descriptive practices for public services staff
- Evaluate archival collection management systems and determine a plan for implementing an ACMS for use in Briscoe Center.
- If approved, plan and guide migration of legacy collection data management systems and tools to the recommended archival collection management system

Major components of the ACMS Pilot Project include:

- Non-encoded archival description (finding aid masters directory)
- Legacy encoded archival description (TARO EAD)
- Installation and maintenance of ArchivesSpace
- Evaluation and testing ArchivesSpace as an ACMS for the Briscoe Center

### **Previous Work/Service to Campus Community:**

The work completed in the Briscoe Center ACMS Pilot project built on workflow analysis and system evaluation of AccesstoMemory (AtoM), which was undergone in Summer/Fall of 2014 for the interdepartmental Digital Preservation and Archival Description Working Groups. At that time, ArchivesSpace was considered beta, and not yet a viable enterprise-level collection management tool. The AtoM system evaluation began with an analysis of our current archival processing workflow which included an inventory of existing tools used in each of the 4 stages of our existing processing workflow (pre-processing, processing, finding aid creation and EAD markup), mapping of our existing processing workflow to steps or functionality in AtoM and time trials to estimate the processing efficiencies gained using AtoM

Work artifacts from the AtoM evaluation and ACMS pilot include proposal for internal advocacy, evaluation matrices, templates for mapping institutional processing workflow, cost projections, functional requirements, and sampling strategy. All of these work artifacts will be presented and distributed to the campus community for reuse and ultimately aid in campus consortial efforts to streamline and standardize archival descriptive practices across University of Texas repositories.

### **Local network of expertise**

During the course of this pilot, the ADTF has been able to identify a local network of expertise within the University of Texas at Austin that has proven to be fundamental to our research. This network serves as a source of peer-to-peer mentorship and information sharing as we towards ACMS implementation.

Members of this network include:

- Metadata coordinator at the University of Texas Libraries, Melanie C. Cofield, has served as a liaison with the University Libraries cataloguing department providing feedback on the appropriateness of ArchivesSpace metadata model to the University of Texas cataloguing environment.
- TARO Steering Committee members working on campus Amy Bowman (Dolph Briscoe Center for American History), Jennifer Hecker (UT Austin Libraries ), Jessica Meyerson (Dolph Briscoe Center for American History) and Donna Coates (Alexander Architectural Archive) have furnished contextual information regarding TARO's ongoing developments relevant to the ACMS evaluation phase.
- Minnie Rangel (IT Architecture and Strategy-Software Developer, University of Texas Libraries), has served as a facilitator to the DBCAH data located on the TARO directory.
- Perry Thompson and Fred Gilmore (System Administrators, University of Texas Libraries) have installed, hosted and maintained sandboxes for testing AtoM and ArchivesSpace

## Recommendations

Based on the key findings of the pilot project and with the hope that the Briscoe Center may begin to systematically approach and streamline our collection data management practices, the Archival Description Task Force (ADTF) makes the following recommendations:

1. We recommend that the Briscoe Center move forward with the planning and implementation of an archival collection management system.
2. We recommend that the Briscoe Center invest in a cost-sharing partnership with other archival repositories on campus to support an institutional ArchivesSpace membership.
3. We recommend approval for a dedicated percent time for members of the ADTF to lead areas of the implementation roadmap identified in this report, and for all members of the research and collections staff to assist with some aspects of data normalization, migration and training activities.
4. We recommend that Briscoe Center bids for contract for encoding services to address our backlog of non-encoded archival description.
5. We recommend that an iSchool graduate student intern be recruited and hired to assist with ACMS implementation activities (including data normalization on encoded EAD and quality assurance on batches of encoded EAD returned by the vendor) and that they be supervised by a member of the ADTF.

## Components of the Project in More Detail:

### ARCHIVESPACE EVALUATION

#### Research

Over the past year, the ADTF, as part of the interdepartmental Archival Description Working Group, conducted several interviews and discussions with peer institutions that have adopted or are in the process of adopting ArchivesSpace as their ACMS (See [Appendix A](#) for details of all the institutions contacted). The majority of institutions contacted were transitioning from Archivist's Toolkit. Our conversations with Rice University have been especially relevant, since we share the same EAD encoding requirements as fellow TARO members. Conversations with institutions implementing AtoM revealed that most of them were adopting an ACMS for the first time, which provided reference points for legacy collection data management and migration projects.

Inspired by the the January 2009 report on archival management tools written by Lisa Spiro for the Council on Library and Information Resources (CLIR), the ADTF developed a functional requirements matrix (see [Appendix B](#)) and evaluated ArchivesSpace and AtoM according to functional priorities (an early analysis of ArchivesSpace functional priorities can be consulted on [Appendix C](#)). More recently,

members of the ADWG attended SAA's *Archival Collections Management Systems* [DAS] webinar in March 2015, in order to check our current work against the most recent findings regarding evaluation and selection of an archival collections management system.

The ArchivesSpace membership model was found to be one of the major pain points during the evaluation. While AtoM provides open access to all their documentation, ArchivesSpace restricts their manual and training materials to members only. During the past months, the Collections Management Tools Roundtable at the SAA began hosting members' unofficial ACMS documentation, including that of ArchivesSpace, that includes workflows, best practices and policies. So far contributors include Arizona State University, Oklahoma State University, University of Nebraska at Omaha, University of Rochester and the University of Southern California. This initiative is just one example of ongoing efforts to maintain a community-driven approach to the development of ArchivesSpace beyond the membership status. Concurrently, there is an ArchivesSpace Google Group that exists as a technical forum for users, and an ArchivesSpace wiki that serves as official communication tool and portal for the community to share their developments and plug-ins. It is also important to acknowledge the efforts of specific institutions including Yale University, who has documented their transition from Archivist Toolkit to ArchivesSpace via project blogs, and who has taken a leading role in contracting with software development companies to extend ArchivesSpace functionality for the rest of the archival community.

Finally, the ADTF has kept an attentive eye to the new descriptive and encoding environment subsequent to the implementation of EAD3 - evaluating the implications of announced changes on the Center's adoption of an ACMS. The report prepared by the SAA EAD3 Study Group in May 2015 regarding software developers attitudes toward EAD3 implementation reveals ArchivesSpace's intention to support EAD3 as soon as possible after the public distribution of the completed schema, offering it as an option for both import and export features. Furthermore, new developments and deployments on EAD3 make this new schema more compatible with ArchivesSpace data model than the current EAD 2002 so that archivist can expect a smoother flow of the data.

## **Methods**

The evaluation of ArchivesSpace included two components: 1) testing of ArchivesSpace migration tools and identification of data cleaning needs; and 2) analyzing the impact of an ArchivesSpace implementation on archival processing and description workflows at the Center. Migration of Accession and EAD records were tested separately, since each type of record has distinct data clean-up needs.

BCAH accessions are managed with a FileMaker Pro database consisting of 29 datafields. This database has been in use since the mid-1980's and currently contains approximately 9000 accession records. Using

the Accession CSV map provided by ArchivesSpace as a starting point, we tested the mapping of our own FileMaker fields. [Appendix D](#) shows a BCAH Accessions to ArchivesSpace data Map that also records additional data such as required fields and controlled vocabularies. Fields have been color-coded based on expected degree of difficulty to import. Blue fields will map easily with a minimal amount of rework. Pink fields will present some extra work, including parsing of data and creation of user-defined fields. Red fields present the data that the ArchivesSpace importer is not designed to handle. More discussion will be required on how to deal with these red fields.

49 accession records were selected to test how well our current accession record fields map to the ArchivesSpace accession record fields. Refer to [Appendix E](#) for a list of the accession records imported as tabular data. Sampling criteria included the acquisition year (a record from each year, 28 in total) and a sample of 21 hand-picked records. All the errors we encountered, as well as the steps we took to resolve those errors can be found in the Accessions error log (see [Appendix F](#).)

A similar workflow was followed for the EAD import testing. Our sample size consisted of 87 EAD records (2.5% of our current 3,567 legacy EADs). Our sampling strategy was based on two criteria. Half of the EADs were hand-picked by the archives processing staff based on their expertise in local archival descriptive practices and encoding use cases (see [Appendix G](#)). The other half were systematically selected according to the results of the XQuery data analysis performed as part of this pilot (see [Appendix H](#)). Any error given by the system during the import process and the solutions tested were recorded on an EAD import error log (see [Appendix I](#)). Once our data was successfully imported, we tested the mapping of our EAD template elements into ArchivesSpace fields using ArchivesSpace's *EAD Import/Export Map* as a reference tool. The [Appendix J](#) shows a BCAH EAD to ArchivesSpace map.

The ArchivesSpace export feature was also tested. Exported EADs files were analyzed in Oxygen and compared to our EAD template in order to identify customization needs. A sample of the exported MARC records have been routed to the University of Texas Libraries staff in order to assure its compatibility with our OCLC and Sierra cataloguing systems.

In order to measure the impact of ArchivesSpace implementation on processing and descriptive workflows, we revisited documentation developed for the evaluation of AtoM during the Summer/Fall of 2014, using the same method to map the tasks in our current workflow to functionalities or features in ArchivesSpace. Special attention has been given to the evaluation of ArchivesSpace user and group permissions. Consult [Appendix K](#) for the ArchivesSpace permissions guide.

## Findings

Considering the outcomes of the testing of ArchivesSpace import functionalities, it is safe to say that both legacy accession data and legacy EADs will require both programmatic and manual cleanup before import. Issues identified during the accession data import include locations, relationships between accessions, and relationships between accessions and resources. While ArchivesSpace supports explicit linking between accessions and resources, the system does not currently support csv import of relationship data. We are aware that accessioning features will require further testing. However,

preliminary findings suggest that user-defined fields may allow us to migrate accession record fields that are unique to the Briscoe Center accession record database.

Other import issues are primarily associated with mapping values stored in our current accession record fields to ArchivesSpace's more granular data model. For example, the "Address" field in our accession record database has to be parsed out and mapped to several address-related fields in ArchivesSpace, including: "Street address," "City," and "Postal code". Likewise, ArchivesSpace requires additional qualifying data that we do not track in our current accession record database. For example, in AS, the "date" field requires a "date label" qualifier. The parsing/discretization of our accession record data and the addition of qualifiers are both tasks that can be done programmatically as an initial data cleanup step. Further batch data cleaning and transformation includes:

- Re-formatting of date fields to ArchivesSpace YYYY-MM-DD date format
- Mapping to boolean values (1/0)
- Mapping of data content to controlled vocabularies
- Cleanup of non utf-8 characters

There are several fields that will require manual data transformation, including collection restrictions. Our database currently stores collection restrictions as boolean values and do not differentiate between user restrictions and access restrictions. Restriction types will have to be determined for each record and mapped to the corresponding field on the migration csv before importing. A complete list of import issues regarding the Accessions can be consulted on the [Appendix L](#).

On import, many of our existing EAD records were found to have significant data incompatibilities with the ArchivesSpace EAD template. Normalizing our DTD-compliant-only EAD records to validate against the ArchivesSpace schema/data model represents one of the main challenges of implementing a collection data management system - it also represents the best argument for it: data consolidation and normalization make it easier 1) to upgrade and migrate to different systems as technologies change and 2) for users to engage with your digital resources in a reliable (i.e., trustworthy) way.

In attempting to understand the scope of the data normalization task, we have already identified a consistent set of data compatibility/data validation issues which can be consulted in [Appendix M](#). The most problematic EAD elements include <extent>, <eadid>, <unitid>, <unitdate>, <container> and <c>. For example, ArchivesSpace requires an <extent> element within the <physdesc> element, while our EAD templates use @extent as an attribute of <physdesc>. ArchivesSpace is also very specific about the data content that it expects for the <extent> element; it must be an integer followed by a space and the physical unit. Any characters before the number (e.g. circa), the use of a period for decimals (e.g. 5.2 ft.), or the use of fractions for small collections (e.g. ½ in.) also cause a validation error. Some of these data incompatibilities can be addressed programmatically using a transformation script. However other incompatibilities, based on inconsistent encoding practices or errors, have been harder to identify suggesting that manual review of some portion of the data will be required.

While the result of the mapping between our EAD elements and ArchivesSpace fields has been mostly a success, we have several cases where there is no one-to-one mapping. This includes the "physical unit"

subfield, which ArchivesSpace automatically parses from the <extent> element, grabbing the string of characters following the extent number (e.g. [3] ft., 5 in.). Similar use cases might require a programmatic parsing of each one of the physical units into its own <extent> element. More relevant, however, is the analysis of the different use cases regarding the collections' restrictions. As a common practice, restrictions have been indicated within the <unititle> or <container> elements on the box level. Restriction information is imported as a string of characters but the boolean value that indicates the system that the collection is restricted does not get activated. Since we consider this a relevant collection management feature, we recommend further testing and analysis of this issue to develop a proper solution.

Overall, one of the most significant findings of this pilot has been the estimated positive impact that ArchivesSpace implementation will have on the BCAH archival processing and description procedures. The mapping of our current workflow to the ArchivesSpace functionalities ([Appendix O](#) and [Appendix P](#)) reveal that ArchivesSpace will facilitate creation of higher quality -- more strictly controlled -- collection data as well as more effective collections management. The implications of implementation include:

- Reshaping of our current processing workflow to accommodate ArchivesSpace functionalities
- Further analysis of how the new ACMS system will impact all areas and departments within the organization
- Development and delivery of training to all the involved staff

## RECOMMENDATIONS

### Tools

- Yale has developed a series of ArchivesSpace plugins available to the users community. We recommend further testing in order to determine which plug-ins will meet the needs of the Center. Examples of plugin functionality includes:
  - Payment plug-in allows to supplies the payment information to the accessions module
  - Materials type plugin allows to keep track of digital materials and other unusual material types
  - Search plugin improves the somewhat limited search option in ArchivesSpace
  - Locations plugin allows for the assignment of a barcode to the containers
- We recommend the development of a validation tool, that may need to be systematically refined, to batch validate EADs previously to its import to ArchivesSpace. Yale staff have been working on a similar tool using schematron, a language for making assertions about patterns found in XML documents. Schematron can be implemented and updated with Oxygen, the XML editor software we currently use to encode our EADs.
- OpenRefine to clean up and parse the Accessions database
- XSLT to cleanup and transform EAD data

### Staff and Time

We estimate that the import of the Accessions records will take around 1 month of FTE. Accessions cleanup requires someone with knowledge of OpenRefine. In addition to programmatic data parsing and cleanup, some manual cleanup will be required. In addition, we may need to develop a solution to import data not included on the ArchivesSpace CSV import tool (e.g. relations, locations). This may include looking into the database underlying ArchivesSpace. Therefore the assigned person should have some knowledge of SQL. QA would be needed following the import to assure data quality and consistency. The tasks can most likely be divided between 2 staff working part time on the project, so that while one is working on the script and data cleanup, the other one is developing a solution to import data not included on the import tool.

The import of the EAD records will also take around 2 months of FTE. The staff profile should include some knowledge of XML, XSLT and Schematron languages. Staff members should also be familiar with XML editors since it will be necessary to develop an ArchivesSpace validation tool and a transformation stylesheet to ArchivesSpace data model. Some time would have to be invested in developing the best strategy to isolate all the cleanup needs that cannot be identified programmatically. QA would be needed following the import to assure data quality and consistency. Tasks can be divided between 2 staff members working part time in the project, so that one person is in charge of developing the transformation and validation tools, while the other person does the testing and reports with feedback for the refinement of the tools.

ArchivesSpace implementation, including the testing of existing plugins, customization/development of plugins and customization of ArchivesSpace taxonomies, is estimated to take around 1 month of FTE. Tasks can be divided between two staff members with different profiles working part time in the project. The installation, customization and development of plugins will require knowledge of JavaScript and familiarity with an AJAX environment. The testing of the existing plugins and customization of taxonomies will require knowledge of the DBCAH archival processing workflow, as well as the DBCAH archival description best practices.

The development of training materials will be an ongoing activity that will keep on evolving along with any decisions taken during the implementation process. An initial investment of 2 weeks of 0.5 FTE can be expected to document the new ArchivesSpace workflow. Once all the decisions have been made regarding ArchivesSpace implementation and the adaptation of the DBCAH processing manual to the ArchivesSpace environment, another month of 0.5 FTE would be needed to document the new processing manual and develop training materials for the rest of the DBCAH staff. Delivery of training can take place on the course of 3 weeks, including group and one-to-one training sessions.

### **Legacy Encoded Archival Description**

In order to analyze our legacy EAD, we surveyed various tools that could be used to work with a large collection of XML documents. In the end, we decided to move forward using BaseX, an XML database that has a built in XPath/XQuery processor. We adopted BaseX as our tool of choice based on its



GUI-based XQuery builder, which allows the user to look for patterns in descriptive practice within any single EAD element.

Using a copy of our Finding Aid masters, we were able to create what BaseX calls a “Collection” made up of our EAD finding aids (some 3,567 XML documents at the time of the copy creation). We pointed BaseX to our copy of the finding aids, stored locally in a folder. This folder became our database for BaseX and any query that we wrote could then be told to run over the entire database of finding aids and return results.

Our discussions led us to believe that the most useful fields to look at in our legacy EAD would be tags related to the title and dates of the collections. For this reason we ran queries constructed to look at the <titleproper> and <creation> tags in our finding aids. Over the course of our investigation we also ran queries to look at the status of our index terms as well as our revision history in the <revisiondesc> tag. Our queries were formatted to output the data into a .csv file so that we could view the results more easily outside of BaseX. This output requirement meant that our queries had to be constructed in certain ways depending on the information we were asking in return. In some cases, a table could be created as the output which would allow for an easy import into a spreadsheet. These queries could then have column headers and some more complex formatting.

## **Findings**

Going into our analysis, we hoped to identify any changes in styles or conventions over times, as well as identify key areas of our data that would require a significant amount of normalization. We found there was no gradual change over time in descriptive practice based on the fields that we analyzed. Rather, we noticed a few large batches of internally consistent descriptive practice that correspond to periods in which a large volume of finding aids were encoded by a single source (be it a third party contractor or a processing archivist).

However, there were still variations in the EAD encoding that became evident during our analysis. The <creation> tag experienced a large variety of styles over time, most dramatically in the <date> tag nested within it. In most cases, the format of Month YYYY is used in the <date> tag, however there is a large variation of formats that have been used in our EAD, the most common being:

- Month DD, YYYY
- YYYYMMDD
- MM/DD/YYYY

Our analysis also revealed the fact that in some cases various EAD fields have not been encoded, or encoded incorrectly. In some cases, our <eadid> tag for our TARO ID has been filled in with a duplicate ID from another finding aid, or simply has #'s acting as placeholders for the ID. These types of errors can be seen in the <creation> and <processinfo> tags.

These variations in style and format create the need for data normalization and clean-up. The normalization of many fields, including dates, will probably need to be done by hand... given the large numbers of variations that have been found in our EAD. Fields that experience a similar variety of formats will also need to be normalized. In the event that there is a field that is consistently formatted incorrectly, that field could be programmatically changed to the correct format.

## **RECOMMENDATIONS**

### Tools

BaseX is not the only lightweight XML database with built-in XQuery functionality. Tools like Zorba and oXygen have XQuery functionality. However, BaseX does have the benefit of a GUI and an easy-to-use XQuery editor that allows you to perform your queries in real-time against a database of XML files. For this reason, BaseX is a great tool that saves you time and effort with your EAD analysis.

Although these queries are easily run against EAD finding aids, there are some interesting quirks that may arise during analysis. Depending on the versions of BaseX or any other tools that you are running, the output of your queries could vary slightly. EAD records are often highly nested, complex files and, as a result, XQuery may try to preserve indents from the XML, resulting in odd-looking outputs. The content is all there, but you may have to look more closely at the results. Best practice is to always double check the XML finding aid when you doubt any output from query.

### Time

You need to spend time familiarizing yourself with the XQuery language as well as the tool that you decide to adopt to help you in your analysis. Once you feel comfortable writing simple queries, you need to spend time looking over the results and verifying their validity against the XML finding aids.

### Staff

It would be best to assign two people to this task initially. One person should be dedicated to the XQuery side of the task while the other focuses on the EAD finding aids and their contents. With an understanding of what the EAD finding aids contain and what needs to be analyzed, that person can then dictate what needs to be found with a query, leaving the query construction to the XQuery individual.

This saves time in the sense that one person does not have to simultaneously learn a query language and digest the existing body of EAD finding aids.

Once query results are obtained the two staff can then discuss results, as well as provide two sets of eyes looking over the results. This helps ensure that as many variations and differences in style in the XML can be identified.

### **FINDING AID MASTERS DIRECTORY**

A Finding Aid Masters (FAM) directory project was developed and implemented to get a grasp on the extent, organization and characterization of our entire holdings of collections at the Center. Our overarching goal was to ensure that each collection has a folder on the FAM directory that contains only the most up-to-date version of its Finding Aid, and that any other material (old or voided inventories; descriptions, or finding aids; memos to the holding record or separation sheets; donor or administrative information, e.g.) is held in a separate, Digital Holding Record (DHR) folder. This would add to the efficiency of EAD outsourcing and help us determine which collections or finding aids need attention by isolating the finding aids from other forms of collection description.

There are a set of logical steps to cleaning up this FAM folder, which have been documented for training and other purposes. The basics are as follows:

- Create a parent folder for each collection title, according to our standardized naming structure, within a Letter Folder (A, B, C, etc.) on the FAM Surrogate folder on the server
- Put the latest version of the finding aid for that collection in that folder, and rename that file to incorporate the naming structure
- Create a parent folder for any collection that has HR material and name it according to same naming structure
- Create a Google spreadsheet for all collections in the Letter folder according to our FAM project template, and fill out all columns in the spreadsheet:
  - Collection name, TARO link (where available)
  - Whether it's a FA, a HR, or a HR without a FA
  - Length and template of the FA
  - Whether it needs to be flagged for review
  - Whether it's part of a specific collection like Photographs or Natchez Trace
  - Any pertinent notes

After a few weeks of staff members completing these Google spreadsheets, we noticed some patterns. We averaged a completion of between 27 and 33 collections an hour, including folder creation and spreadsheet documentation. Thus far, 1,831 out of an approximate 6,821 total collections have been surveyed and completed in spreadsheet form. Approximately 62% of the 1,090 collections surveyed, currently have TARO records. However this result is skewed by the disproportionate number of TARO records created by volunteers for the A subdirectory between 2007 and 2008. We cannot assume that a similar percentage of the remaining collections will have TARO records.

We identified seven different templates distinct templates that have been used at the Center, two of which we have merged into one category we call “Fix”. This category includes finding aids in unreadable form, with IBM special characters taking the place of letters and numbers, as well as finding aids with only a collection title and location, but no Accession Number or description of any kind. Out of 1,813 collections surveyed, 88 finding aids fall into this category.

The other templates are as follows:

- New – BCAH logo: the most up-to-date versions of the finding aids we could find, these include correct formatting and the latest version of the BCAH logo in the top left corner. 512 finding aids.
- New – old CAH logo: most are pretty recent, and some have correct formatting, but some do not. These contain the former Center for American History logo in the top left corner. 104 finding aids.
- TARO PDF: in these instances we chose to download a PDF version of the TARO record if it was the most up to date version of the finding aid or if no other finding aid was available. 612 finding aids.
- Old – Paragraph: these are old, unformatted finding aids which have paragraph formatting for description, but columns for index terms. 327 finding aids.
- Old – Columns: these are old finding aids formatted in such a way that all descriptive information can be found under bold headed columns. 297 finding aids.

A characterization of the makeup of the collections surveyed that we deemed important pertains to which collections have final finding aids, which have digital holding record folders, which have both, and which are simply pointer files to other collections. Thus far we have noticed that nearly  $\frac{3}{4}$  of the total collections surveyed, 1,314 collections, contain finding aids but no digital holding record content. One hundred and forty collections contain only a holding record folder but no final finding aid, 320 have both, and there are 39 pointer files, which make references to material within other collections.

Out of the 1,813 collections that we surveyed, 496 have been flagged as needing review. These contain a variety of issues, not limited to:

- Unprocessed collections or accessions
- Multiple accessions that are not described in one final finding aid
- Collections that are currently restricted but should be reviewed to be made accessible
- Collections with multiple TARO records, whether that be one record in multiple parts or duplicate or varying versions of records
- Collections in which the FAM does not match the TARO record
- Collections with inadequate finding aids (lacking inventory, description, etc.) that should be reviewed for research value and described or possibly deaccessioned
- Collections that need to be merged into other collections
- Misspellings of creator names and collection titles, including some with duplicate entries under the misspellings

We additionally noted 445 collections with finding aids of a length shorter than one half of one page. In most of these cases, the finding aids are lacking adequate description and so these might want to be set aside for review as well.

It is recommended that the Finding Aid Masters cleanup project be allowed to continue at a rate of 5 hours per week of one full-time archivist's schedule. At this rate of work, if the average rate of cleanup continues, the remaining 5,008 collections will be able to be surveyed in fewer than 40 weeks, or 10 months. We would also request, during this period, that a supervisor dedicate approximately 5 hours per month to reviewing necessary collections and making recommendations for action, which the archivist could fit into her workflow.

## **Review of Findings and Recommendations**

### **FINDINGS**

- We have a large backlog of legacy data in multiple formats and standards
- Both legacy accession data (currently in FMP database) and legacy EAD will require both programmatic and manual cleanup before import
- ArchivesSpace will provide workflow efficiencies, eliminate versioning issues and provide added functionalities (reporting, collection management) that we haven't had access to before

### **RECOMMENDATIONS**

1. We recommend that the Briscoe Center move forward with the planning and implementation of an archival collection management system.
2. We recommend that the Briscoe Center invest in a cost-sharing partnership with other archival repositories on campus to support an institutional ArchivesSpace membership.
3. We recommend approval for a dedicated percent time for members of the ADTF to lead in the implementation, and for all members of the research and collections staff to assist with some aspects of data normalization, migration and training activities.
4. We recommend that Briscoe Center bids for contract for encoding services to address our backlog of non-encoded archival description.