This study is a survey of how archival institutions have addressed retrospective conversion of their legacy finding aids to EAD. Legacy finding aids are defined as those that exist on any format other than EAD, but are primarily paper or document/text files. The survey was issued online and asked 20 questions regarding background data and approaches to conversion, including staffing, software, funding, and lessons learned from the process. The findings indicate that many institutions began the process in 2008 or later and are still in the process of completion. Institutions use multiple combinations of staff and software, seldom use encoding vendors, and often are not grant-funded projects.
Table of Contents

Table of Contents .................................................................................................................. 1
Introduction .......................................................................................................................... 2
Literature Review .................................................................................................................. 4
Methodology ......................................................................................................................... 16
Findings ............................................................................................................................... 19
Analysis ............................................................................................................................... 24
Conclusion ............................................................................................................................ 29
Suggestions for Future Studies ............................................................................................. 30
Bibliography .......................................................................................................................... 32
Appendix ............................................................................................................................... 34
  1.1 Appendix 1: Listserv Invitation for Survey ................................................................. 34
  1.2 Appendix 2: Survey ....................................................................................................... 34

Figure 1: Institution Type ...................................................................................................... 20
Figure 2: Other Staff Who Worked on Legacy Finding Aids .................................................. 21
Figure 3: Legacy Finding Aid Formats .................................................................................. 22
Figure 4: Year Recon of Legacy Finding Aids Began ............................................................ 22
Figure 5: Software Used ....................................................................................................... 23
Introduction

Finding aids have long been the established means of content description for processed archival collections. They have endured style and format changes, but the core information on a collection’s provenance, administrative information, historical or biographical background, and scope and content remains constant. The dramatic increases in the retention of paper documents since World War II has made it nearly impossible for modern researchers to find relevant material without the use of finding aids. They were essential for researchers to assess a collection from outside their immediate area, albeit more difficult to obtain prior to the widespread use of the Internet.

As archives and researchers have adapted their practices to incorporate computers and the Internet, finding aids have also migrated online. Through the establishment of content standards such as *Describing Archives: A Content Standard* (DACS) and encoding standards such as *Encoded Archival Description* (EAD), many archives began the shift from producing paper finding aids to digital ones.

While archives increasingly produce new finding aids as online files, there remains the task of also converting all previously existing finding aids. These finding aids will be referred to as legacy finding aids, and the process of converting them will be referred to as retrospective conversion (recon). Legacy finding aids can take the form of paper, text/document files, HTML code, spreadsheet, card file, or combination thereof.
Many institutions still have unconverted legacy finding aids, usually due to a combination of staff shortages, budget restraints, and technological barriers (Yaco, 2008).

This study will examine the strategies used by institutions that have already completed, or are in the process of completing, recon of their legacy finding aids. Strategies include funding, staff, process, technology, and changes to the finding aid and/or collections. The study will focus only on finding aids converted to EAD.

The purpose of this study is to gather updated data specifically on institutions that have addressed recon of their legacy finding aids. Much of the current scholarship focuses on adoption and implementation of EAD, with less attention to legacy finding aids. Many times, implementation of EAD is primarily focused on new finding aids, not previously existing ones. There are 90 globally located institutions listed on the SAA (Society of American Archivists) EAD Implementor’s webpage, which leads one to assume that there are many other institutions who have yet to address their legacy finding aids, or have not reported on how they did. The data collected in this paper has several practical applications. It may provide institutions that have not yet converted their legacy finding aids with insight on how to proceed. It should also provide current data for researchers analyzing EAD-related archival trends.

The audience for this paper is primarily working professionals in the archival field. The secondary audience is library and information science students, and this paper could potentially be used as a source in a research paper or thesis on a related topic. Related papers are published in journals like *American Archivist, Archivaria*, and *Journal of Archival Organization*, and working papers produced by organizations such as OCLC.
It may also have appeal to those in the broad field of information science as EAD uses XML. However, this is not a technical paper.

There are three primary research questions to be answered. The first is what kinds of institutions have completed recon of their legacy finding aids. This includes type of institution, staff size, and funding. The second question is how did the institution develop their conversion plan? This includes questions related to time, software used, and changes made to the finding aids and the collections. The third question is what would the institution do differently if it could start the project over. This addresses lessons learned from the process.

The sample consists chiefly of academic, government, and historic institutions, with varying staff sizes and collection holdings. They were contacted for the survey via three listservs: the Archives and Archivists (A&A) listserv maintained by SAA, SAA’s Description listserv, and the EAD listserv maintained by the Library of Congress (LC).

**Literature Review**

The literature regarding EAD has remained consistent since its inception in the 1990s. The scholarship fits into four general categories: analysis of its development, academic discussions about its role within the profession, practical and how-to articles on implementation, and case studies and surveys about implementation. Much of this literature addresses the barriers to implementation such as technological complexity, funding, staff, and time. The archives community has worked hard to address these barriers, yet many institutions continue to face difficulties with EAD.

This information becomes extremely important to consider when repositories begin to assess how they will approach recon of their legacy finding aids. There is a lot to
consider when developing a plan for conversion to EAD. Will it be a specific project? Is there extra funding available? What formats are the legacy finding aids in? Are there plans to rewrite or reprocess the collections prior to conversion? Will the work be outsourced? This is not even taking into account all the technological issues with EAD. In short, one must do one’s research before committing to a plan.

The articles selected for this literature review are representative of the scholarship on EAD. The predominant theme is assessments of the state of EAD adoption throughout the time it has been in use. There are few previously published articles and studies specifically focusing on recon of legacy finding aids. Therefore, the majority of literature cited addresses the wider issue of EAD adoption. This literature is still relevant as recon of legacy finding aids does get addressed, albeit often in brief terms. Additionally, institutional relationships and attitudes toward EAD manifest and reveal trends in implementation.

Bouche’s case study of Yale is an early examination of the implementation process. Tatem and Pitti’s pieces represent early concerns about EAD and its future within the profession, with Tatem focusing more on the negative and Pitti taking up the defensive. Marshall’s study on implementation sets the template for future studies by Yakel & Kim, and Yaco. Prom and Wisser both survey implementation from the technological angle, while Redding and Goulet and Maftei examine content revision and standardization. Finally, the OCLC report published in 2010 by Combs, Matienzo, Proffitt, and Spiro update the scholarship with current best practices for implementation.

Yale was one of the first major institutions to implement EAD. Bouche’s case study examines the methods used for recon of its legacy finding aids. The case study
finds that implementation of EAD provides both technological hurdles as well as intellectual and content ones (Bouche, 1997).

The case study is primarily about lessons learned in the process. Many early problems were due to compatibility issues with software, such as advertised functions from vendors that didn’t work the way they were supposed to. SGML was also not widely supported among the leading web browsers, creating additional compatibility problems. Routine and workflow also had to be altered to accommodate EAD (Bouche, 1997).

Ultimately, this early project was considered a success, and provided a template for other institutions to emulate. The biggest lesson learned, Bouche finds, is that priorities must be defined. This includes how much time and effort to devote to legacy finding aids versus the creation of new ones (Bouche, 1997). On this point, it is clear that many institutions in the following years pursued a little of both, setting up the backlogs of legacy finding aids still awaiting conversion.

Tatem sums up the early concerns of many archivists regarding EAD in a 1998 paper. She examines the advantages and disadvantages of implementing EAD finding aids within the framework of Rogers’s innovation diffusion theory, a theory that other archivists will use in subsequent analyses. Her analysis relays the cautious approach archivists seem to take with new technologies, which tends to express concern about EAD rather than outright dismissal. As she explains in her paper, all her findings should be considered tentative, as EAD was still relatively new in 1988 (Tatem, 1998).

Tatem acknowledges the advantages of using EAD, namely unmediated remote access, non-proprietary standards that should endure technological shifts, and incentive
for grant funding. She does express concerns as well. Complexity, and the steep learning curve required for implementation are some of the challenges institutions will face when using EAD. Her chief concern, however, is the archival profession’s lack of understanding of users’ needs; this will hinder both the design of digital finding aids, and the measured effectiveness of them (Tatem, 1998). She offers no specific solutions, but in her assessment archivists will not make full use of EAD until they have a clearer understanding about how finding aids will function online.

In 1999, Pitti presented to the archival community his rationale for implementing EAD as the standard for archival description. As technology continued to advance at a rapid pace, archivists were keen to take advantage of these developments, but also had concerns about durability and obsolescence. Additionally, the structure of finding aids represent collections comprised of material organized at various levels (collection, series, subseries, file, item) not suitable for MARC, which is not a hierarchical format. EAD is flexible enough to work with these different levels, along with additional matter such as provenance, extent, biography, and scope and content. EAD was also developed to be an encoding standard independent of hardware and software that could endure changes. He also saw standardizing encoding for finding aids as the means for creating universal union access through databases and indexes allowing researchers to more easily access more collections worldwide (Pitti, 1999).

Furthermore, Pitti established the need for standardization of the components of archival description. Included in this is the need for a content standard, providing specifics on the elements of a finding aid and how they should be written. Lastly, EAD
should have communication standards to enable easy communication from computer to users (Pitti, 1999).

Pitti, who was involved with the development of EAD, established the reasons why archivists should adopt and implement EAD. He was cognizant that EAD was in continual development, and that while many institutions had already adopted it, many more had not. He fails to address barriers to implementation, a topic that numerous archivists will address in the first half of the 21st century.

Marshall’s 2002 exploratory study on EAD implementation in repositories serves as the model for studies in the years to come. At the time of her study, EAD was still relatively new, and Marshall indicates that few studies like hers had been conducted. She breaks down the literature at the time of her study into 3 categories: general articles about the development of EAD, technical and how-to articles geared at those considering implementing EAD, and case studies of EAD implementation in repositories. This breakdown of the categories of EAD literature remains relevant all the way to the present. Much like Tatem before her, Marshall uses Rogers’s theory of diffusion as it relates to organizations as a means of assessing the transition from innovation to institutionalization (Marshall, 2002).

She devised a questionnaire to gather data regarding general information about the repositories’ EAD program, staffing and funding for EAD projects, and future plans for EAD. Specifically addressing recon of paper finding aids, the survey asks if EAD implementation was for retrospective conversion or the creation of new finding aids, and what criteria was used for selection of finding aids to encode. The assumptions for the survey were that for institutions that implemented EAD, but did not incorporate it into
their larger descriptive program, there would be little evidence of the institutionalizing of the standard. Additionally, administrative actions such as funding and staffing would differ in institutions that institutionalized EAD rather than those who were just experimenting with it (Marshall, 2002).

The survey was sent to 34 institutions via email and received 16 responses. The institutions chosen for the survey were selected from the SAA EAD Roundtable website. Her findings indicate that institutional satisfaction with EAD, and that some repositories have routinized the standard. Her initial hypothesis regarding institutionalization was also proved incorrect, with 63% of respondents building EAD into their operating costs, instead of just relying on grants. Regarding legacy finding aids, the study found that 75% used EAD for new and legacy finding aids, although some did not convert all of them, but converted those with higher usage statistics. At least one institution decided to rewrite its legacy finding aids in order to improve them, which also slowed the process down (Marshall, 2002).

The importance of Marshall’s study is that it establishes an early attempt to study different aspects of EAD implementation in institutions. It relays that recon of legacy finding aids was a priority for many institutions in the formative years of EAD implementation. It also demonstrates the difficulty of locating and gaining the cooperation of institutions for this kind of study.

In 2002, Prom conducted a usability study and survey for the *EAD Cookbook* as a means of assessing EAD implementation for smaller institutions. Developed by Michael Fox in 2000, *EAD Cookbook* is a middleware package that combines implementation documentation, downloadable template files, and downloadable XSLT stylesheets. The
idea was to provide archivists with a simplified means of learning and implementing EAD at their institution.

Prom’s methodology was a survey and usability study of users of *EAD Cookbook*. His sample of 27 individuals were selected from an invitation sent out on the EAD listserv (presumably this is the listserv maintained by the Library of Congress) as well as a group of archivists identified to Gina Minks of University of Tulsa Special Collections as users of *EAD Cookbook* (Prom, 2002).

Prom’s survey looked at institutional data, how *EAD Cookbook* has been used, time spent using it, user attitudes about it, and the importance placed on EAD implementation at the institution. The survey was aimed at smaller institutions, and 19 of 27 respondents worked at colleges or universities. Nearly all respondents stated that EAD implementation was a high priority, yet only 12 had EAD finding aids online at that time. 23 respondents had 25 or fewer finding aids encoded. While all respondents found that *EAD Cookbook* simplified implementation, technological barriers still existed. Prom found *EAD Cookbook* to be effective, but also limited. His overall assessment echoes the findings of Tatem and others: until the learning curve and technological expertise barrier is lowered, archivists will continue to struggle with EAD implementation. Middleware like *EAD Cookbook* is a step in the right direction, even if it’s not a complete panacea (Prom, 2002).

Redding’s 2002 study examines archival practices regarding revision of finding aids when converting them to EAD. He puts forth that too much attention has been paid to the technological aspects of EAD implementation, and not enough to intellectual control. In particular, he is interested in the adoption of content standards during recon of
legacy finding aids. This study was conducted prior to the release of *Describing Content: An Archives Standard* (DACS), but it does mention ISAD(G)\(^1\). The survey asked 53 institutions about their EAD implementation strategies in regards to revision of finding aids and implementation of content standards (Redding, 2002).

The results of Redding’s study found that there was increasing adoption of standardization, but not uniformly so. This leads to inconsistencies in content validation, as well as not fulfilling the purposes of EAD. He further argues that archivists need to shift focus from presentational qualities to data-centric ones (Redding, 2002). DACS largely addresses this issue, and is an approved standard by SAA. Redding’s survey is indicative of a crossroads moment in the adoption of EAD in institutions: the embracement of new technology without fully realizing how to integrate it intellectually with the content.

Yakel & Kim’s 2005 study of EAD adoption builds on the works of Marshall, Roth, and Minks and Curtis, and also frames their scholarship around Rogers’ theory of diffusion of innovation. Yakel and Kim used survey methodology, and a population comprised of 399 institutions that sent staff to EAD workshops form 1993-2002. Their response rate was 34%, or 135 people. Of the 135 responses, 58% of them had not yet adopted EAD (Yakel & Kim, 2005).

Their survey found that staff and collection sizes played a significant role in EAD adoption. Possible reasons for this include more funding and time to invest in developing the skill sets needed to work with EAD. They also found this to be true regardless of whether or not the institution was involved with a consortia, a concept meant to assist

---

\(^1\) International set of descriptive archival standards. This is an international archival content standard, and DACS is the U.S. version of this standard.
smaller archives in achieving parity with larger ones. They found a link between archives
who previously adopted MARC and those who adopted EAD. Perhaps their most
significant find was that the diffusion of knowledge regarding EAD did not correlate with
adoption, meaning that practices such as outsourcing proved a viable means of
implementation. However, institutional knowledge of EAD remains lower in such
circumstances. Technological compatibility and levels of expertise continued to manifest
as concerns regarding implementation. With regard to legacy finding aids, they only
acknowledge the potentially costly retrospective conversion of older data, and that
implementation of EAD may be contingent on funds to do so (Yakel & Kim, 2005).

While Prom’s survey focused on a single means of EAD implementation, Wisser,
on behalf of the EAD Working Group of the Society of American Archivists, conducted a
usability survey of EAD tools in 2005. Wisser’s survey identified the commonly used
tools used by archivist and usability of the tools in 4 categories: authoring, publishing,
discovery, and knowledge. She also makes recommendations based on the results of the
survey, and provides an extensive overview of the tools and best practice guidelines
available to archivists. Her methodology was a survey conducted online that received 70
responses. The survey questions were a mixture of open-ended and close-ended questions
(Wisser, 2005).

Wisser found that 41 of the 70 respondents had used *EAD Cookbook*, and the
majority held a positive opinion of the toolkit. Many used EAD authoring software like
NoteTab, XMetal, XMLSpy, and Oxygen. Common needs of the respondents were
stylesheets, templates, and documentation on best practices. Aside from tool needs,
respondents need more time, more staff dedicated to EAD, and ways to proofread EAD
encoded finding aids. Wisser’s specific recommendations as a result of the survey include an update to Michael Fox’s (1997) suggestions to reflect EAD’s transition from a SGML standard to XML, a tools portal run by those in the EAD community, and more research into the effectiveness of EAD for users (Wisser, 2005).

In another study from 2005, Goulet and Maftei examined the recon of 40 legacy finding aids dating from 1863-2000 at the Archives Départementales des Pyrénées-Atlantiques. Many of these finding aids lacked modern archival standards of structure and did not have clean typeface, making it not possible to OCR (optical character recognition) them, thus requiring retyping. It was found that the recon of these finding aids provided an opportunity to update and rewrite them for a modern audience (Goulet and Maftei, 2005)

The project brought forth several difficulties with recon of legacy finding aids. Limits in time prevented the revision of introductions to conform standards. Plans were made, however, to revisit these in the future. Footnotes also needed to be integrated into the finding aid. To accomplish this, they added the information to tags as needed. Finally, indexes proved to be difficult to address digitally. They were too time-consuming to encode, and also lacked standardization (Goulet and Maftei, 2005).

The study reveals that with recon of legacy finding aids, the content needs to be preserved while allowing for changes in structure, format, and standards. This may increase the funds and time needed for the process, but will bring a consistency to the finding aids across the collections. Echoing Redding, they support the need for standardized content, in this case ISAD(G). Goulet and Maftei conclude that archivists
should be able use both the standardized content and the structure of EAD to present information to novice users (Goulet and Maftei, 2005).

A more recent survey study about EAD implementation is by Yaco in 2008. Her study focuses on the problems institutions have faced when implementing EAD. She used a primarily closed question survey and a sample of 16 archivists answering an electronic survey, along with telephone interviews. Her findings support Yakel & Kim’s conclusion that staff size makes a difference in the implementation process, with 12 of her 16 respondents employing fewer than 2 FTE archivists. Yaco found that barriers to EAD implementation also included a lack of time, planning, and institutional support. Technological issues, such as the complexity of EAD, also proved to be a barrier consistent with previous studies. Regarding legacy finding aids, 10 respondents want to rewrite or update legacy finding aids prior to conversion, likely causing additional time and funding constraints (Yaco, 2008).

Her findings establish three primary barriers to implementation: lack of staff, the middleware gap in delivering EAD finding aids to the Internet, and plans to rewrite legacy finding aids prior to conversion. Her solutions focus on striking a balance between outsourcing work as funding will allow while simultaneously incorporating EAD into the larger description program. Finally, Yaco concludes that the archival profession needs increased training in technological skills in order for archivists to keep pace with changes (Yaco, 2008).

In 2010, OCLC issued a report that addresses the political, logistical, and technological barriers to EAD implementation that have been discussed in the preceding literature and others up to the present. The report, authored by Combs, Matienzo, Proffitt,
and Spiro, assumes the reader has a basic understanding of what EAD is, but is unsure of how to formulate a plan for implementation.

Information in the first section addresses political and logical barriers to implementation. This includes clear definitions of what EAD is, and why it is important for archival description, talking points for management, and plans for how to begin implementation. They also recommend taking into account Greene and Meisner’s “More Product, Less Product” concept when implementing EAD, such as just starting out with collection level finding aids. With regard to legacy finding aids, the report advocates that institutions not rewrite the finding aids prior to conversion. Their argument is that even if the descriptions are outdated, they still can provide access to the collection, and this should be the top priority of the repository (Combs, Matienzo, Proffitt, & Spiro, 2010).

The second section addresses technological issues. This includes data migration, commercial vs. open-source software selection, and publishing. Much like Wisser’s 2005 survey, the appendix includes lists of consortia and EAD aggregators, software, templates, forms, XSLT stylesheets, XML publishing platforms and more. The report echoes the findings of previous studies in that there is no single path to EAD implementation. It suggests that institutions make informed decisions, document and adhere to them. Creation of guidelines for how they will proceed with EAD will help ensure best practices, and above all else, keep the encoding simple (Combs et al., 2010).

The OCLC report is a very practical document. It is presented in simple language, with a tone both sympathetic to the difficulties of EAD, but also optimistic about its merit. It is largely the culmination of the working knowledge of EAD that has been studied, surveyed, and refined over the past 13 or so years. Despite documents such as
this, institutions still struggle with implementing EAD as well as utilizing its potential for efforts such as retrospective conversion of legacy finding aids.

The literature on EAD implementation reveals the archival community’s willingness to adopt and adapt to EAD, despite myriad problems along the way. Along with changing technology and implementation strategies, institutions also recognize the importance of migrating their old files to match the format and standardized content of the new. It would appear that while the complexities that Bouche encounters with Yale may have subsided, new and different challenges have manifested, evidenced by Yaco and the OCLC report. Implementation of EAD is still an ongoing project for many institutions, and recon of legacy finding aids will continue to be a part of that process.

**Methodology**

The methodology used for this study was an online survey. It was determined that this would be the most efficient and effective tool for gathering data quickly. They are easy to distribute, cost-effective, and convenient for respondents. Online surveys are also an effective tool for acquiring direct answers to questions related by topic variables from a sample in multiple geographic areas (Wildemuth, 2009). Finally, the collected data does not require data-entry thus the turnaround for analysis is much shorter (Selm and Jankowski, 2006).

While an online survey has been determined the most appropriate means of gathering data, the disadvantages should be acknowledged. Privacy must be ensured, creating a need to de-identify data and ensure secure storage. Random samples can be
difficult to obtain. It is also challenging to anticipate response rates. Online surveys have the potential to have repeat responses, or multiple responses from the same institution. Finally, online surveys have the potential to exclude segments of the population that are not technologically proficient at using the Internet, or do not have regular access to it (Selm and Jankowski, 2006).

This study used an unrestricted sample directed at institutions that have already completed, or are in the process of completing, recon of their legacy finding aids. An unrestricted sample is defined as allowing anyone with access to the link to take the survey. While this was the case for this study, it was only posted on archives-related listservs germane to the target sample. A larger group of respondents would reduce the sampling error, but aside from membership in SAA, the total sample population cannot be calculated (Selm and Jankowski, 2006).

Several measures were enacted to ensure the privacy of the respondents. The survey did not ask for any identifying information, such as name, email, or institution name. Each respondent is assigned a randomly generated string of numbers and letters that serves as their unique identifier. The software does collect the IP addresses of the respondents, but this information will not be used or retained.

Participants for the survey were recruited online via three listservs that cater to archival topics and EAD. The first is the Archives & Archivists (A&A) listserv hosted by SAA. This is an open forum for current topics in archival practice and theory. As it is a general forum, it is also likely to have the largest number of viewers. The second listserv is the Description listserv, also hosted by SAA. This is a closed forum only available to SAA members. Its specific focus is archival description, therefore a good candidate for
the survey. The third listserv is the EAD listserv maintained by the Library of Congress. This forum is EAD specific, and also a good source of potential respondents. Please see Appendix 1 for the post used for the survey invitation.

The survey was conducted using Qualtrics, a research software suite. The software was selected for its ease of use in survey design, distribution, and data analysis. Additionally, Qualtrics offers added levels of privacy and security that other free survey programs like Surveymonkey cannot match.

Whether developed on software or on paper, survey design remains largely the same. The survey for this paper was planned out on paper and then modified in Qualtrics. In order to facilitate higher response rates, the survey adhered to the design principles suggested by Wildemuth. These principles include only asking necessary questions, using neutral language, avoiding double-barreled questions, and allowing respondents to skip questions they do not want to answer. The questions were primarily objective, requiring non-scalar responses appropriate for questions that don’t rely on personal attitudes and opinions. However, two open-ended questions at the end of the survey provided the opportunity for respondents to provide additional comments regarding the topic. Lastly, the survey was kept to a minimum length (Wildemuth, 2009).

The questions were designed to gather a general overview of information about the institutions in regard to staff, holdings, and how recon of their legacy finding aids has been addressed. The first 3 questions gather information on institution type, number of full-time staff who work with archival materials, and size of holdings. Questions 4-6 address the number of legacy finding aids that have been converted to EAD, along with those that have not, as well as format. Questions 7-8 cover the timeframe for conversion.
Questions 9-14 ask about staff who directly work on recon of legacy finding aids, grant funding, and encoding vendors. Questions 15-17 inquire about reprocessing and re-engineering of collections and legacy finding aids, as well as software. Finally, questions 18-20 provide opportunities for respondents to answer open-ended questions regarding additional comments and lessons learned about what they would have done differently. Please see Appendix 2 for the complete survey.

There are limits to this study. There was the risk of not receiving enough respondents to make reasonable claims for the field based on the gathered data. The questions asked take a broad approach to the topic. Those interested in a specific aspect of recon of legacy finding aids, such as budget costs for a specific project will not find that information in this study. The only ethical concern with this study is anonymity. Since the survey is web-based, and not targeting potential respondents via email, no personal names, institution names, or email addresses were acquired. The only identifying data that existed was the IP address, and this information was not used or retained.

This study gathered quantitative data. The majority of the questions are closed, many with yes/no or have unranked selections of possible answers. Data analysis was done with the tools included with Qualtrics. The software was also used to generate all charts used to visualize the data.

Findings

The survey was released to all three listservs on February 14, 2011, and data was collected through March 17th. The sample consists of 86 respondents to the survey. It is important to note that not every respondent answered every question.
59 of the respondents identified their institution as academic or a special collections. 7 were historic societies, 8 were state or federal government institutions, and 11 were identified as “other.” The write-ins for “other” included public libraries, religious institutions, non-profits, art gallery, and a museum. No respondents identified as a corporate archive.

<table>
<thead>
<tr>
<th>Answer</th>
<th>Response</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>State or federal government</td>
<td>8</td>
<td>9%</td>
</tr>
<tr>
<td>Academic or special collections</td>
<td>59</td>
<td>69%</td>
</tr>
<tr>
<td>Historical society</td>
<td>7</td>
<td>8%</td>
</tr>
<tr>
<td>Corporate</td>
<td>0</td>
<td>0%</td>
</tr>
<tr>
<td>Other</td>
<td>11</td>
<td>13%</td>
</tr>
<tr>
<td>Total</td>
<td>85</td>
<td>100%</td>
</tr>
</tbody>
</table>

Regarding staff, 70 respondents had 10 or less full-time employees who worked directly with archival collections, 10 had between 11 and 22 full-time employees, and 3 institutions had 30 or more. Only 7 respondents had hired a full-time staff member specifically for recon of legacy finding aids.

40 respondents had 1 full-time staff member working on recon, and 20 had 2. 17 institutions had between 4 and 7 full-time staff members work on recon. The highest number of full-time staff working on recon was 12, and only 1 respondent answered this way.

All respondents in the sample had used other types of employees to work on recon of their legacy finding aids. The other types of employees used were graduate students
(32), interns (25), para-professionals (24), undergraduate students (22), and volunteers (17). Most institutions used more than 1 kind of other employee.

**Figure 2: Other Staff Who Worked on Legacy Finding Aids**

<table>
<thead>
<tr>
<th>#</th>
<th>Answer</th>
<th>Response</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Para-professional staff</td>
<td></td>
<td>24</td>
</tr>
<tr>
<td>2</td>
<td>Undergraduate students</td>
<td></td>
<td>22</td>
</tr>
<tr>
<td>3</td>
<td>Graduate students</td>
<td></td>
<td>32</td>
</tr>
<tr>
<td>4</td>
<td>Interns</td>
<td></td>
<td>25</td>
</tr>
<tr>
<td>5</td>
<td>Volunteers</td>
<td></td>
<td>17</td>
</tr>
</tbody>
</table>

There was a wide variety in the response to the number of legacy finding aids already converted to EAD. The largest number was 750,000, and this belonged to a government archive. 8 respondents had converted 1,000 or more finding aids, 38 had converted between 100 and 1,000, 24 had converted less than 100, and 12 had yet to begin the conversion process. Regarding the number of legacy finding aids yet to be converted, 14 still had over 1,000, including one with 750,000 (the same government archive previously mentioned) and another with 18,000. 34 respondents had 100 to 1,000 left, and 20 had less than 100. Only 10 respondents answered with 0.

Regarding file formats of legacy finding aids, paper (69) and document/text files (67) were the most common. 17 respondents also selected the “other” category, and majority of these answers reflected some kind of database.
Nearly half of the sample, 33 respondents, began recon of their legacy finding aids in 2008 or later, with 3 beginning the process in 2011. 33 began the process between 2000 and 2007, and 7 began prior to 2000. 81% (63 respondents) of the sample indicated that they have not yet completed converting their legacy finding aids.

Regarding the use of grant funds, only 22 respondents had received grant funding for recon of legacy finding aids, with 60 stating they did not. Similarly, only 13 had used an encoding vendor, with 67 stating they did not. Of those who responded that they had
used an encoding vendor, 3 indicated that the vendor converted all of their legacy finding aids to EAD.

Two questions addressed changes made to the collections or legacy finding aids prior to recon. The respondents were almost evenly split on re-processing collections prior, with 38 who did and 40 who did not. The majority, 69 respondents, re-engineered or updated their legacy finding aids to modern archival standards prior to recon.

There was a wide variety in the EAD authoring software used. Respondents could select more than one option. Oxygen had the highest with 27 users, XMetaL followed with 25, Archivists’ Toolkit had 20, NoteTab had 18, Archon had 9, EAD Cookbook had 6, and 4 had used XMLSpy. 24 respondents, however, also selected “other” and write-ins included in-house software, MS Access and Excel, NotePad++, OhioLINK EAD FACTORy, Dreamweaver, and Xiasoft.

<table>
<thead>
<tr>
<th>#</th>
<th>Answer</th>
<th>Response</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>NoteTab</td>
<td>18</td>
<td>24%</td>
</tr>
<tr>
<td>2</td>
<td>Oxygen</td>
<td>27</td>
<td>36%</td>
</tr>
<tr>
<td>3</td>
<td>XMetaL</td>
<td>25</td>
<td>34%</td>
</tr>
<tr>
<td>4</td>
<td>XMLSpy</td>
<td>4</td>
<td>5%</td>
</tr>
<tr>
<td>5</td>
<td>EAD Cookbook</td>
<td>6</td>
<td>8%</td>
</tr>
<tr>
<td>6</td>
<td>Archon</td>
<td>9</td>
<td>12%</td>
</tr>
<tr>
<td>7</td>
<td>Archivists’ Toolkit</td>
<td>20</td>
<td>27%</td>
</tr>
<tr>
<td>8</td>
<td>Other</td>
<td>24</td>
<td>32%</td>
</tr>
</tbody>
</table>

Finally, when asked if the respondent could restart the conversion process would they proceed in the same fashion, 44 said yes and 30 said no. Write-in statements for those who said no and the write-in statements for additional comments will be addressed in the Analysis section.
Analysis

The survey exceeded my expectations in terms of response rate. Previous surveys, although often emailed directly to institutions, have not always received high response rates. Yakel & Kim had a response rate of 34%, Wisser received 70 responses to her online survey, and Yaco’s sample consisted of 13 completed surveys and 3 telephone interviews. This survey had 86 valid respondents. If anything, the response rate indicates that the recon of legacy finding aids is still an ongoing concern for many institutions.

The demographics of the respondents were mostly academic or special collections, and 70 indicated that they had 10 or less FTE staff who worked directly with archival collections. No institutions identified themselves as corporate archives, which is unsurprising since most function for internal use rather than for public research. It is likely, therefore, that there is significantly less need for EAD.

The data suggests that a large number of archival institutions still have unconverted legacy finding aids. Of the 81 respondents reporting the number of remaining legacy finding aids, only 10 had 0. 48 had between 100 and 750,000 remaining. 63 also reported that they had not yet completed recon. Additionally, nearly half of institutions surveyed did not begin recon of legacy finding aids until 2008 or later. There weren’t many respondents who directly addressed this issue, but one answered: “We should probably have been earlier adopters of EAD, but needed to wait until there was a low/no cost alternative. We had some philosophical questions about EAD (why bother?) that prevented us from adopting it.” It is probable that other institutions had similar sentiments.
In terms of who is actually working on the recon of legacy finding aids, it seems to usually be 1-2 FTE staff, along with the help of graduate students, undergraduates, interns, volunteers, and para-professional staff. The write-in answers reveal a number of staffing approaches. One stated “We re-assigned 3 full-time catalogers to work on EAD conversion up to 50% of their time.” Another respondent discussed the project in terms of hiring a project archivist: “We have approval to hire a temporary, part-time archivist to help convert finding aids to EAD. This project will commence in March 2011. We also plan to acquire Archon to aid in the conversion process as well as provide a front end for displaying and searching our finding aids. Depending on technical issues, this is expected to coincide with hiring a project archivist.”

The work is also being done primarily in-house, with only 13 respondents opting to use an encoding vendor for any part of the process. For institutions that did use encoding vendors, experiences varied. One respondent’s solution for improving output with an encoding vendor was to create a template and tagging guide: “The outsourcing worked well because we put together a very detailed tagging guide and provided sample finding aids marked up by hand so the vendor was clear on our needs. We also maintained frequent communication with the vendor so that as questions came up they were answered promptly.” However, another respondent illustrates some of the problems institutions can encounter as well:

Most finding aids were outsourced to a vendor with the intent of importing them into Archivists' Toolkit. The remainder were manually entered into Toolkit by students and para-professional staff. The outsourcing portion was grant-funded and the grant was written so that we could outsource all of our finding aids (about 6700 pages) at once. This approach sounded great to administration but was hard on the staff responsible for (minor) re-processing, DACS compliance and markup. The pressure of completing all 6700 pages within the grant period led to mistakes.
While we technically had 100% of our finding aids in EAD by August 2010, we are still individually reviewing the output and correcting misplaced tags.

There was no single EAD authoring software that dominated use. Much like in Wisser’s 2005 study, data indicates that institutions often used more than tool. Most of the tools she surveyed, such as XMetal, NoteTab, and Oxygen, are still actively being used at present (Wisser, 2005). There were not many comments specifically regarding software experience, but one respondent mentioned a favorable experience with Archon, and another expressed displeasure with Archivists’ Toolkit.

Anecdotal evidence from the write-in answers suggests that recon of legacy finding aids is not a simple, straightforward process. There are three constraints mentioned that were persistent throughout the comments: low priority, time, and complications with the collection and/or legacy finding aid.

Several respondents indicated that recon of legacy findings aid was a low priority, something worked on sparingly, or delegated to graduate students. One respondent made it clear that primarily only new finding aids were being created with EAD. Another stated:

Although 2 full-time employees have worked on converting legacy finding aids, that task is not their primary job responsibilities. At this point, encoding legacy finding aids is not a priority as the process is incredibly time consuming; rarely entails simply encoding work. Instead, the finding aid is cross-checked against the collection with inevitable issues in the inventory. Often rehousing needs to occur as the collection was processed long ago. Often series or entire collections need reprocessing.

Time is probably the most significant factor complicating recon of legacy finding aids. This respondent summed it up succinctly:
People have a hard time finding the time to spend on working out the details of the conversion. It takes some concentrated effort that staff doesn't feel they have the time for. Everyone has too many projects to be able to focus on converting legacy finding aids. However I am supervising 5 students that are helping me to update guides (including incomplete EAD guides). This is a hard chore because of all of the decisions that need to be made along the way.

Another indicated that despite spending one-fourth of their work week on EAD, the process of addressing legacy finding aids has been very slow:

I spend approx. 25% of my work week on some aspect of EAD. The older the finding aid the more updating must be done to it before I can convert it. I had a volunteer student help with our 2 largest finding aids which probably saved me at least 1 year's worth of work. I started conversion as a pilot project for 1 year. I hope to complete the rest of the finding aids in the next 5 years, perhaps it will take less time than that.

Many of the comments revealed the complicated nature of recon of legacy finding aids. This includes employing numerous strategies to complete the work. Several responses illustrate this point, such as this answer:

Legacy finding aids come in many forms and we have employed many strategies and technologies to encode them—OCR, Spreadsheets, XSL, and brute force re-keying. We have been converting legacy finding aids as we find time without any dedicated staff. The most challenging and time consuming part of the project has not been the actual encoding of legacy finding aids, but locating such finding aids and making sure they describe the complete collection. The process often involves gathering up all of the descriptive tools for a collections and combining them into one coherent and standards-compliant finding aid.

Another discusses using workarounds:

Most of our work has been done using in-house workarounds. For example, the container lists were converted using saved find-and-replace queries authored in Dreamweaver; the resulting code was then pasted into XML documents and validated. Similarly, older finding aids have been updated to more current standards using global find-and-replace queries. We have also decided not to recode indexes, but rather to post .pdf surrogates using the <otherfindaid> tag. Personally, I've found tools like AT and Archon unacceptably restrictive and so have used other formats to generate valid FAs in considerably less time and then imported them.
The last question asked what respondents would have done differently. Responses from this question indicate that there were many “lessons learned” from these projects. Several stated that they wished they had hired more staff, or allocated more staff resources to the project. One respondent mentions: “If we had the luxury, I would hire a full-time or part-time staff member to manage the conversion project. So far, other priorities and job responsibilities have prevented us from completing the project.” Another echoes this sentiment: “I believe I would make the case for a short-term full-time staff position to either do the conversion/processing or at least free up the existing staff to concentrate on that by using the short-term position for reference duties, etc.”

Several respondents said they would have altered how the project was carried out. One aspect mentioned was the need to find ways to automate or batch-process the project. Another aspect was asserting more control on the project, including student workers and encoding vendors. One respondent mentions:

We would outsource in smaller batches and pay more attention to quality control on the front end. We would probably limit outsourcing to the larger finding aids, since we found that a student worker can enter a 20 page finding aid in Archivists' Toolkit in about the same amount of time it takes a staff member to prepare the same finding aid for outsourcing (check DACS compliance, prepare encoding worksheet, write up tagging instructions, scan paper documents).

Another states: “I worked with four grad students and I would probably retain tighter oversight of their work processes.”

This respondent voices questions about the justifications and long-term durability of the project:

I would select the same finding aids but would ask better questions before agreeing to participate in the larger grant-funded online consortia which has been plagued with lack of infrastructure support and adequate funding. Its long term
sustainability is questionable. The staff time tied up in meetings, travel to workshops, and consortia governance has come at a cost which has been a challenge to justify to Board members and administrators. EAD has become a kind of litmus test for gauging stewardship, professional competency, technical knowledge and repository credibility/parity. We participated in the project for those reasons. Our collections were online and discoverable long before conversion to EAD.

Finally, several respondents address reprocessing and prioritizing. One respondent felt that reprocessing slowed the process down too much: “Staff training, learning PHP, and probably not doing as much reprocessing as we did would have made it more efficient.” Others wished they had re-appraised their collections beforehand: “We would re-appraise collections and determine whether or not the level of detail they were processed to was appropriate.” Another would have altered the workflow this way: “In an ideal world, I would prioritize the finding aids and do the most important first. Given more staff, I would create a workflow where the finding aids would be updated to meet current standards, the information on the finding aids would be checked for accuracy and then converted to EAD.”

**Conclusion**

The goal of this study was to gather current data on how institutions are addressing recon of legacy finding aids. The results indicate that while many institutions have begun to address this, it is often a slow and complicated process that is not completed in a any short amount of time. The data also shows that there is no single strategy that will work for every institution.

Because there is such diversity among archival institutions, it is difficult to make grand sweeping statements about strategies. However, based on the survey results, there
are a few general assessments that can be made. It appears that many of the same problems that are present when implementing EAD continue with recon of legacy finding aids: lack of time, staff, and funds, as well as finding the appropriate technological solution. Additionally, legacy finding aids exist on multiple formats, and it is likely that institutions have used multiple software programs to address this.

The results also indicate that recon of legacy finding aids are rarely a top priority for institutions. Rather, it seems like institutions fit it in when they are able to. While several indicated that used encoding vendors or grant funding, the majority did not. Finally, a major factor in the slowness of recon of legacy finding aids is that many of them are out of date, and need to be re-written as they are encoded. Concurrently, nearly half the respondents determined that the collections needed re-processing prior to recon.

**Suggestions for Future Studies**

This study was designed to be a broad overview of how archival institutions are addressing the conversion of their legacy finding aids to EAD. Future studies related to this topic could benefit from a more granular focus on the topic. A survey that focused specifically on grant funded recon projects would likely be of use to institutions that are planning on going this route. A survey with this focus could also potentially compile the results into a list of recommendations or template that institutions could reference in planning.

Consortium-based recon projects are a topic not addressed in this survey, although several respondents made mention of them in the open-ended questions. A future survey could incorporate questions regarding consortiums, or focus on them.
Finally, a future survey on the topic could focus more directly on how archivists are addressing the complications to the process. This could include more detailed questions on reprocessing collections, re-writing legacy finding aids, and establishing priorities and timelines. Additionally, future surveys could gather more information on how to address the different formats of legacy finding aids.
Bibliography


Appendix

1.1 Appendix 1: Listserv Invitation for Survey

I am a graduate student at the University of North Carolina at Chapel Hill in the School of Information and Library Science. I am currently writing my master’s paper on strategies for conversion of legacy paper finding aids to EAD. Legacy finding aids have been defined for this project as those that only exist on paper, document/text file, HTML code, spreadsheet, card file, or combination thereof.

I am conducting an online survey of the strategies used by institutions to address their legacy finding aids, and am currently seeking participants. The survey is anonymous, and no identifiable data will be retained to ensure privacy. If you are willing to fill out the survey on behalf of your institution please click the following link:

https://uncodum.qualtrics.com/SE/?SID=SV_6x3FWPRdNtCjGSM

Thank you for your time, and I greatly appreciate those who are willing to participate. Please feel free to email me with any questions or concerns regarding the survey.

Thank you,

Adam Fielding
University of North Carolina at Chapel Hill
afieldin@email.unc.edu

1.2 Appendix 2: Survey

Online Consent Form: University of North Carolina-Chapel Hill Consent to Participate in a Research Study Adult Participants Social Behavioral Form

IRB Study #_11-0086____________________ Consent Form Version
Date: 1/25/11______________ Title of Study: Strategies for Retrospective Conversion of Legacy Finding Aids to EAD Principal Investigator: Adam Fielding UNC-Chapel Hill Department: School of Information and Library Science UNC-Chapel Hill Phone number: 202-390-8604 Email Address: afieldin@email.unc.edu Co-Investigators: Faculty Advisor: Jackie Dean Funding Source and/or Sponsor: Study Contact telephone number: 202-390-8604 Study Contact email: afieldin@email.unc.edu

What are some general things you should know about research studies? You are being asked to take part in a research study. To join the study is voluntary. You may refuse to join, or you may withdraw your consent to be in the study, for any reason, without penalty.
Research studies are designed to obtain new knowledge. This new information may help people in the future. You may not receive any direct benefit from being in the research study. There also may be risks to being in research studies. Details about this study are discussed below. It is important that you understand this information so that you can make an informed choice about being in this research study. You will be given a copy of this consent form. You should ask the researchers named above, or staff members who may assist them, any questions you have about this study at any time.

What is the purpose of this study? The purpose of this research study is to learn about the strategies institutions have applied to conversion of their legacy finding aids to Encoded Archival Description (EAD). Legacy finding aids are defined as those that only exist on paper, text file, HTML code, spreadsheet, card file, or a combination thereof.

Are there any reasons you should not be in this study? You should not be in this study if you do not work an archival institution or have no knowledge of how your current or past institution converted their legacy finding aids to EAD.

How long will your part in this study last? The survey should take approximately 15-20 minutes, and there are no follow up questions.

What will happen if you take part in the study? This study consists of a survey with 2 questions. The questions are multiple choice, yes/no, and short-answer. You may skip a question at any time.

What are the possible benefits from being in this study? Research is designed to benefit society by gaining new knowledge. You will not benefit personally from being in this research study.

What are the possible risks or discomforts involved from being in this study? There are no known risks or discomforts for this study.

How will your privacy be protected? No personal or institutional names will be collected in this survey. Any identifiers such as IP addresses will be deleted prior to data analysis. The data will be kept in a password-protected drive. Participants will not be identified in any report or publication about this study. Although every effort will be made to keep research records private, there may be times when federal or state law requires the disclosure of such records, including personal information. This is very unlikely, but if disclosure is ever required, UNC-Chapel Hill will take steps allowable by law to protect the privacy of personal information. In some cases, your information in this research study could be reviewed by representatives of the University, research sponsors, or government agencies for purposes such as quality control or safety.

What if you want to stop before your part in the study is complete? You can withdraw from this study at any time.

Will you receive anything for being in this study? You will not receive anything for taking part in this study.

What if you have questions about this study? You have the right to ask, and have answered, any questions you may have about this research. If you have questions, complaints, concerns, or if a research-related injury occurs, you should contact the researchers listed on the first page of this form.

What if you have questions about your rights as a research participant? All research on human volunteers is reviewed by a committee that works to protect your rights and welfare. If you have questions or concerns about your rights as a research subject, or if you would like to obtain information or offer input, you may contact the Institutional Review Board at 919-966-3113 or by email to IRB_subjects@unc.edu.

Title of Study: Strategies for Retrospective Conversion of Legacy Finding Aids to EAD
Principal Investigator: Adam Fielding
Participant’s Agreement: I have read the information provided above. I have asked all the questions I have at this time. I
voluntarily agree to participate in this research study. Please click the button below to agree to terms and begin the survey
☐ Disagree
☐ Agree

How would you describe your institution?
☐ State or federal government
☐ Academic or special collections
☐ Historical society
☐ Corporate
☐ Other ____________________

Approximately how many full-time staff work with archival materials?

What are the approximate total holdings of your institution? (in linear or cubic feet)

To date, approximately how many legacy finding aids in your institution have been converted to EAD?

Approximately how many legacy finding aids still need to be converted?

What format were the legacy finding aids in prior to EAD conversion? (check all that apply)
☐ Paper
☐ Document/text file (.DOC files, etc.)
☐ HTML code
☐ Other ____________________
☐ Spreadsheet
☐ Card files

When did your institution begin its conversion of legacy finding aids to EAD?
When did your institution complete its conversion of legacy finding aids to EAD?

- 1996
- 1997
- 1998
- 1999
- 2000
- 2001
- 2002
- 2003
- 2004
- 2005
- 2006
- 2007
- 2008
- 2009
- 2010
- 2011
- Not yet completed
Has your institution hired full-time staff dedicated to the conversion of legacy finding aids?
☐ Yes
☐ No

Has your institution received grant funding to support legacy finding aid conversion work?
☐ Yes
☐ No

Did your institution use an encoding vendor for any portion of its legacy finding aid conversion?
☐ Yes
☐ No

Approximately what percentage of legacy finding aids were converted by an encoding vendor?
☐ 25%
☐ 50%
☐ 75%
☐ 100%

How many full-time staff directly work or have worked on converting legacy finding aids?

Have any of the following worked on converting legacy finding aids? (check all who apply)
☐ Para-professional staff
☐ Undergraduate students
☐ Graduate students
☐ Interns
☐ Volunteers

Were any of the collections re-processed prior to EAD conversion?
☐ Yes
☐ No
Were any of the legacy finding aids re-engineered or updated to modern archival standards prior or during conversion to EAD?
☐ Yes
☐ No

What EAD authoring software was used? Select all that apply.
☐ NoteTab
☐ Oxygen
☐ XMetaL
☐ XMLSpy
☐ EAD Cookbook
☐ Archon
☐ Archivists' Toolkit
☐ Other ____________________

Do you have any additional comments regarding your institution's approach to converting its legacy finding aids to EAD?

If you could restart the conversion process, would you still do it in the same way?
☐ Yes
☐ No

What would you do differently?