
This study describes a survey of the legacy finding aids in the Southern Historical Collection in the Manuscripts Department at Wilson Library, at the University of North Carolina at Chapel Hill. The survey was conducted to assess the content, scope and quality of the legacy finding aids in preparation for retrospective conversion of those finding aids into Encoded Archival Description (EAD) format.

Most archival repositories are confronted with the problem of converting archival description prepared before the advent of computers and modern archival description into EAD. Because retrospective conversion is a resource and labor intensive task, the surveying of the finding aids to be converted is a critical step that should precede the actual encoding. This paper is a case study of one repository’s retrospective conversion needs and the survey that resulted as part of the planning process as an example of how an archive can go about classifying their legacy finding aid in order to incorporate retrospective conversion into their departmental workflow.

Headings:

Archives--retrospective conversion

Encoded Archival Description (Document Type Definition)

Special Collections--case studies

Historical collections--North Carolina

University of North Carolina at Chapel Hill. Library. Southern Historical Collection
PLANNING FOR RETROSPECTIVE CONVERSION OF LEGACY FINDING AIDS IN THE SOUTHERN HISTORICAL COLLECTION AT THE UNIVERSITY OF NORTH CAROLINA AT CHAPEL HILL: A CASE STUDY

by
Amy Crow

A Master’s paper submitted to the faculty of the School of Information and Library Science of the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Master of Science in Library Science.

Chapel Hill, North Carolina
November 2007

Approved by
Matthew Turi
**Introduction**

Archivists and users of archives have long employed paper-based finding aids—the intellectual intermediaries between the vast quantities of information contained in a collection and the highly superficial amount of information about a collection contained in a catalog record—to gain research access to and control over a repository’s collections. As researchers increasingly use the Internet for a greater share of their preliminary investigation, however, archives are finding it increasingly necessary to provide more access to their materials online. Most archival repositories have some web presence; for example, a site describing the repository and its holdings. To allow online users more access to information about their collections and give archivists the ability to conduct much more rapid searches both in and across collections, many repositories have also begun to provide electronic versions of their finding aids on their sites. Finding aids, as historical documents themselves, exist in myriad formats, despite increasing consensus in the archival community about content standards for archival description.\(^1\) The advent of Encoded Archival Description (EAD), a metadata container language, and *Describing Archives: A Content Standard*, the complimentary manual that dictates (in the loosest of terms) how to fill the containers that EAD provides, have made the presentation of finding aids online more systematic and standardized. All repositories with finding aids produced before 1996, the year the beta version of EAD was released, will eventually have to deal with the issue of legacy finding aids—those finding aids produced before

---

In the spring semester of 2007, I undertook a field experience with the Public Services division of the Manuscripts Department at Wilson Library, at the University of North Carolina at Chapel Hill. Library science students in the Masters Degree program at the School of Information and Library Science at the University of North Carolina at Chapel Hill can have up to two field experiences—or semester-long internships—as part of the curriculum; I chose to do a field experience in the Manuscripts Department as a means of increasing my familiarity with how such a repository works. I had the opportunity to carry out a specific project as my field experience; under the direction of Laura Clark Brown, the head of the Public Services, I conducted a large scale survey of the collection of legacy finding aids in the Southern Historical Collection.

The staff in both Public Services and Technical Services had, for quite some time, been aware that a systematic assessment of the legacy finding aids was necessary in order to begin any concrete planning for the retrospective conversion of those aids into EAD. The Manuscripts Department has been producing finding aids for their collections in Encoded Archival Description, or EAD, format for almost a decade, and approximately one thousand of the collections in the SHC are described online by EAD finding aids. While this represents a significant achievement and certainly sets Manuscripts ahead of many archival repositories in terms of finding aid accessibility, there remains the issue of
the collection of legacy finding aids. The term “legacy” signifies a finding aid produced before the department implemented EAD; many of the legacy finding aids pre-date the use of computers in the archive. There are upwards of 2500 of these finding aids, housed in 137 binders in the Manuscripts Department, and they span almost three quarters of the twentieth century, the earliest having been produced in the 1930s. My semester-long project involved developing and using a survey instrument to collect a set of data about each of these legacy finding aids, entering the survey information into the department’s master database, then running a series of queries against the information collected in order to provide Public Services and Technical Services with a body of data that described the state of the legacy finding aid collection.

**History of Description at the Southern Historical Collection**

The earliest of the finding aids in the SHC collection are those that were produced in the 1930s. The majority of these Depression-era finding aids were produced under the auspices of the Historical Records Survey, a New Deal era Works Projects Administration (WPA) program established in 1935. The stated aim of the Historical Records Survey was to “make available to any interested person a convenient guide to the manuscript collections in the United States.”\(^2\) The North Carolina Project, a state-based division of the Historical Record Survey, was established the following year, 1936, and continued into the 1940s as the North Carolina Historical Records Survey Project. The project aimed for both standardization and accessibility in the finding aids produced under its auspices. Survey employees were provided with pre-printed forms that directed them to record a specific set of information about each archival collection; in North

---

\(^2\) *Guide to the Manuscripts in the Southern Historical Collection of the University of North Carolina* (Chapel Hill: The University of North Carolina Press, 1941), v.
Carolina, the project director instructed his workers to carefully document the types of material in the collection, as well as topics and individuals of significance.  

In the mid-1940s, the archivists at the SHC developed an in-house accession sheet that replaced the less-detailed WPA Historical Records Survey form. The accession sheets had sections for provenance information, date and terms of acquisition, biographical or historical sketches of the creators, dates and regions covered, and the size of the collection. These accession sheets served as de-facto finding aids for small collections; larger collections were often represented by more in-depth surveys, which could include information such as an index of significant proper names, places, institutions, as well as much more granular description. Until the 1980s, the SHC used variations of the accession sheet as a template for their finding aids. In the 1970s, however, the staff began to write brief descriptive guides (in lieu of an actual accession sheet) after the processing of a collection, and more detailed guides were created for the collection depending on necessity and staff availability.  

In the early 1980s, the staff of the SHC revised the format of finding aids for their collections to bring them into alignment with emerging national archival descriptive practices. Finding aids were revised to contain a biographical sketch, a scope and content note, an extent note in cubic feet rather than in item numbers, access restriction and copyright information, shelf lists and container listings. As arrangement practices likewise evolved to make archival materials more accessible, the finding aids at the SHC increasingly reflected series description as the staff began to divide collections without

---


obvious original order into series based on type of material rather than the older practice of blanket chronological ordering. By the 1990s, the size of the collection at the SHC had burgeoned to the point that it was necessary to implement a policy defining different levels of processing for incoming materials: minimal processing, or surveying a collection at a superficial level to provide basic accessibility information and full processing, or detailed arranging and describing of a collection at all levels. This policy allowed for the staff to spend their processing and descriptive efforts on the collections deemed to be of greater research value.\(^5\) It also resulted in finding aids that, while standardized, were vastly different in terms of the amount of archival description presented.

A grant from the National Endowment for the Humanities in the early 1990s provided the SHC with the first great influx of processing resources since the 1930s; this funding underwrote the SHC’s first large-scale retrospective processing project.\(^6\) The retrospective processing work that took place was done under the guises of preservation and preparing of OCLC catalog records to represent the collection in MARC. This project helped to make a significant number of collections more accessible as it resulted not only in MARC records but often in the preparation of new finding aids to reflect new arrangement decisions. With newly word-processed finding aids and catalog records, the Manuscripts Department had the digital raw material necessary to begin uploading collection information to the Internet in 1994; by 1996, 1200 SHC finding aids were available online. In the same year, EAD was introduced in the archival community.

\(^5\) Ibid, 26-30.

\(^6\) Ibid, 30.
Encoded Archival Description is a markup language that was written by archivists to express the specific parameters of archival description. There had been slow movement since the 1970s toward the development of a standardized descriptive language for use in archives, beginning with the work of the National Information Systems Task Force (NISTF) in 1977. This movement was impeded, however, by the archival community’s long-held belief that the unique character of each archival collection necessarily precluded any standardization in favor of unique, localized approaches. An extensive survey conducted in 1979 of a disparate group of repositories revealed, however, that descriptive practices did not vary greatly across archives and manuscript collections. With this realization, NISFT set out to develop an archival description-compliant version of MARC that would enable the electronic capture of finding aids—the USMARC Format for Archival and Manuscript Control (MARC AMC). When archivists began to upload finding aids onto networked servers in the early 1990s, it became clear that MARC AMC was insufficient to contain complete finding aids, and out of this insufficiency came Daniel Pitti’s Berkeley Finding Aid Project and EAD.  

Encoded Archival Description is one of the latest tools that archivists have begun to employ to give users the means of reaching into archival collections without actually being on-site. Paper-based finding aids only serve those users who have access to the physical location in which they are housed. Daniel Pitti, the main architect of EAD, justified its development in these terms: its purpose was to “mitigate the geographic distribution of collections” that necessarily limited the ability of researchers to access

---

7 Steven L. Hensen, “‘NISTF II’ and EAD: The Evolution of Archival Description,” American Archivist 60 (Summer 1997): 286-298.
primary materials. Not only did archives need to respond to the increasingly digital character of research, but archivists needed a means to express finding aids online that would preserve not only the intellectual nature and content of the finding aid, but also the significant hierarchy and relational aspects of archival description that comprise the finding aid. Pitti and his team originally wrote EAD in the form of a Standard Generalized Markup Language (SGML) (in its later incarnation it has been translated into Extensible Markup Language, or XML) Document Type Definition in order that the standard could be hardware and software independent, as well as largely compatible with Internet search structures. The development of EAD thus gave archives a succinct, nationally acceptable, and standardized language with which to make the finding aid accessible online. Implementation of EAD did not require refiguring of archival descriptive practices, as it was specifically built to represent the particular nature of those practices.

The Manuscripts Department at UNC-Chapel Hill developed an EAD implementation program in 1998, and the first encoded finding aids were mounted on the web that same year. By 2001, Technical Services staff members were writing finding aids directly in EAD, rather than in Microsoft Word, finalizing the transition to the encoding standard. Currently, all of the collections in the Southern Historical Collection are represented online in two formats: each collection has a MARC record in

---


10 Laura Knodel, “The Evolution of Archival Description at the Southern Historical Collections,” 34.
the OPAC as well as a dedicated web page on the Manuscripts Department’s web site. More than half of the collections have complete online finding aids on their page, in XML, HTML, or ASCII format; those that do not have an online finding aid are represented by an abstract of the collection, drawn from the MARC record abstract, and a note directing the user to contact reference services for more information. Most of the abstract paragraphs can also be found in hardcopy format in the legacy finding aids, as they were produced as part of a large-scale cataloging and reprocessing grant that the Manuscripts Department received in the early 1990s. While each collection is represented online by some type of electronic surrogate, there has yet to be any large-scale reformatting of the department’s large collection of legacy finding aids to the end of making all SHC finding aids available in EAD and in MARC online.

Since the implementation of the standard in the late 1990s, the Manuscripts Department’s Technical Services staff has done some retrospective conversion—the re-writing of paper-based finding aids into EAD—but conversion has been on an ad hoc, rather than systematic, basis.\textsuperscript{11} Retrospective conversion of a large body of finding aids into EAD is not an inherently straightforward process, as finding aids can take a myriad of formats, not all of which are neatly compatible with EAD. Writing the year after the introduction of EAD, Michael Fox cautioned that implantation of the standard would entail a great deal more than “simply buying computer software and sitting down to mark up finding aids.”\textsuperscript{12} In addition to considering the needs of one’s users, institutional objects, realistic allocation of resources and workflow, Fox noted that not all finding aids

\textsuperscript{11} Interview with Public Services and Technical Services Staff, 8 November 2007.

\textsuperscript{12} Michael Fox, “Implementing Encoded Archival Description: An Overview of Administrative and Technical Considerations,” American Archivist 60 (Summer 1997): 330-343.
would mark up seamlessly into EAD; newly processed collections, he commented, might be “clean” to encode in EAD, however, older finding aids could require significant editorial and stylistic revision.\textsuperscript{13}

**Literature Review**

The architects of EAD knew that the conversion of legacy finding aids would be far more challenging than the development of new EAD-compliant aids, and wrote the EAD document type definition with a great deal of flexibility built in to accommodate the fact that the finding aids that it would be used to encode would be significantly diverse. Despite this flexibility, however, there is little evidence in the literature that archival repositories have undertaken retrospective conversion on a large scale. While a good amount of literature exists on EAD implementation, there is much less about the application of the standard to already existing legacy finding aids.\textsuperscript{14}

The survey that I conducted falls under the scope of the larger retrospective conversion planning project that is being undertaken at Manuscripts; archival retrospective conversion itself is only one type of retrospective conversion in the metadata universe. The term “retrospective conversion” or “recon” was first coined in relation to the migration of traditional card-based catalog records into MARC, and the vast majority of literature on retrospective conversion deals with the creation of electronic catalog records. Understandably, the output of articles on this topic has slowed considerably over the last decade because many libraries and repositories have completed

\textsuperscript{13} ibid.

their catalog conversion projects. The process of converting card catalog records to electronic format differs in significant ways from that of converting finding aids—bibliographic records, unlike archival description, have a much higher degree of standardization, and contain much less information. Recon of catalog records into MARC also occurred on a much more systematic, and ultimately national, basis than archival retrospective conversion; libraries have largely realized the goal of a single integrated system for catalog records with the Online Computer Library Center (OCLC), while a single union database for archival records has yet to be realized.

The literature on card catalog recon is interesting to consider in relation to archival retrospective conversion because the two processes share many of the same overarching goals: improving access to information as well as improving collection management, supporting regional and national resource sharing, and ensuring a means of preserving electronic data. Recon of both card catalogs and finding aids involves some of the same workflow considerations, as well. In both cases decisions about where the conversion work will occur—whether in-house, contracted out, or a combination of both—must be made. Likewise, in both cases, the process of recon is labor and cost extensive to the point that it is usually impossible to go back and redo the work. Advance planning is therefore critical to all retrospective conversion projects. No formal, extensive body of literature on archival recon comparable to that on bibliographic recon

---


yet exists. This lack of published literature is understandable given that the conversion of finding aids is a process that is barely a decade old, and many archival repositories have yet to begin any sort of EAD implementation project.17

The few articles on retrospective conversion that have been published more recently, however, tend to focus on archives or special collections because these types of repositories have materials that require more intensive handling and are less compatible with automated conversion. An article on the evolution of Allegheny College’s catalog notes the unique problems involved with the recon of special collections materials, from inadequate inventories to language barriers.18 Archivists and special librarians at museums have also begun retrospective conversion projects for their catalogs and have published article on the process; an article on recon at the Tate Gallery Library in London details the technical steps involved in moving information from card format to digital format, while an article about recon at the Pulliam Education Center at the Eiteljorg Museum of American Indians and Western Art, in Indiana, provides a more narrative style overview of the decisions and timeline for the conversion project there.19 The overarching purpose for the recon projects at all of these institutions was to provide


remote user access to collection description and, ultimately, to increase usage. Access is one of the underlying goals of all retrospective conversion, whether of card catalogs or finding aids; and the “only actual purpose,” of any mode of access—from MARC record to finding aid—“is to link users and materials,” wrote Beth Whittaker in a recent article. To remain viable in the increasingly digital research environment, archives and special collections are pressed to make more and more aspects of their holdings, from finding aids to the primary materials themselves, accessible, which generally translates to available online.

There is some literature on the wider impact a retrospective conversion project can have on the collections that the records represent. In the late 1990s, the archivists at the American Heritage Center at the University of Wyoming found that the retrospective conversion project that they undertook to improve access to their materials yielded some valuable reappraisal work as well as some new collection management initiatives. While the retrospective conversion of finding aids was not the subject of this project, as it predated all but the earliest implementation of EAD, the process of creating online catalog records allowed archivists at the American Heritage Center to put in place new systems of control over the collections (particularly through organized de-accessioning), an outcome that can result from a finding aid conversion project as well.

Soon after the introduction of EAD, Dennis Meissner reported from the Minnesota Historical Society that their finding aids, while perfectly adequate in the

---


context of in-house use, did not translate well to the online environment where the user would be interacting with the information without the benefit of a trained archivist at hand to explain the terminology and arrangement specific to archival description.

Meissner’s team at the Minnesota Historical Society responded to this problem by developing a new model for their finding aids that would be both EAD compliant and remote user-friendly. Meissner’s article, however, does not address the internal technical problems that retrospective conversion of finding aids raises: vastly differing granularity of description, incomplete information, and non-standardized formatting.

More recently, at the Archives Départementales des Pyrénées-Atlantiques in France, archivists faced the task of converting a body of finding aids produced between 1863 and 2000 into EAD. While much of the re-keying and encoding work was contracted out, the archivists at the Archives Départementales, like those at the Minnesota Historical Society, developed a new model for their finding aids—the conversion project gave them the opportunity to “update finding aids that were in some cases obsolete, and sometimes even to reengineer them completely.”22 This repository chose to merge the finding aids on the basis of subject rather than to respect original collection boundaries described in the legacy finding aids. This retrospective conversion route is very different that the route planned at the Manuscripts Department at UNC-Chapel Hill, but their reflections on how to begin the process of retrospective conversion are relevant because they raise common issues about the importance of advance planning and preparation of the legacy finding aids before beginning any encoding work.

Archivists at the University of California, Berkeley and at Duke University were among the few early implementers of EAD who focused on retrospective conversion as part of their initial implementation project. These two repositories set as their goal the encoding of all of their finding aids, which required planning for how to deal with the legacy finding aids, as well as developing EAD-compliant standards for the creation of new finding aids. Because both Berkeley and Duke were members of a consortium of repositories that were involved in developing a union database for their finding aids, the staff decided to write a set of guidelines for the retrospective conversion of legacy finding aids so that the encoding of these often disparate type documents could be streamlined in such a way to facilitate inclusion in the database. This set of guidelines, entitled “The Encoded Archival Description Retrospective Conversion Guidelines” set forth a list of rules, some of which were mandatory and others of which were suggestions, in the attempt to make the production of consistent retrospectively converted finding aids possible.

Both institutions also conducted surveys of their legacy finding aid collections; the authors noted that “gaining control of [the] collections of finding aids and selecting finding aids for conversion [were] key activities” in the retrospective conversion

---


24 This document was later made available online to assist other repositories in their EAD conversion planning and projects. “The Encoded Archival Description Retrospective Conversion Guidelines: A Supplement to the EAD Tag Library and EAD Guidelines,” available online at [http://sunsite.berkeley.edu/amher/upguide.html](http://sunsite.berkeley.edu/amher/upguide.html).
Tracking databases were used at both institutions, and the general information that was collected at Duke and at Berkeley was as follows: presence of finding aid for the collection; whether the finding aid was in machine readable format; whether the finding aid needed to be scanned or re-keyed; and whether the collection is cataloged. At Duke, they listed their paper legacy finding aids in a Microsoft Access database, and developed a schema for describing the contents of those finding aids. The electronically formatted finding aids were also surveyed and information about the nature of the electronic files was added to the database. They also recorded the number of pages in the finding aid, and whether it appeared to be in a “regular and consistent” format. After completing the survey of the finding aids, the staff at Duke made decisions about how to prioritize their retrospective conversion; frequency of usage was their primary determining factor, followed by ease of conversion.

At Berkeley, the staff involved in planning the retrospective conversion projects set up Microsoft Access databases to track the process and retain vital statistics about the finding aids and the work involved in converting them to EAD. As of 2001, they were also projecting the development of a Web-accessible database set up to hold twenty five sets of information about each finding aid that would allow for multi-institution input of information. In beginning to plan their retrospective conversion project, the staff at the Manuscripts Department at UNC-Chapel Hill is certainly following the same type of process that both Duke and Berkeley undertook; the survey that I conducted of the legacy

---


26 ibid.
finding aids, while tailored to the needs of the Southern Historical Collection, is similar in function to the surveys that archivists at Duke and Berkeley conducted.

The importance of surveying a repository’s collection of finding aids is echoed in some of the literature about the Virginia Heritage Project (VHP), a consortial effort based in Virginia to produce a union database of EAD finding aids. Members of the consortia decided to target collections on African American history and culture in Virginia, and selected over one thousand collections from the involved repositories; to represent these thousand collections online required the encoding of some fifteen thousand pages of finding aids.27 One of the special collections librarians at Old Dominion University, a VHP consortia member, wrote a brief account of her repository’s participation in the project in which she noted that they had completed an information gathering survey designed by the VHP Task Force. “Evaluate your collections and be familiar with the conditions and formats of your finding aids,” she wrote, “this information will be invaluable…”28 While the article does not include any detail about the nature of this survey, it is further evidence of the necessity of some type of survey of the finding aid collection in order to plan and execute an efficient and effective EAD implementation or retrospective conversion project.

Methodology

Like most archival repositories, the Manuscripts Department at UNC-Chapel Hill has a substantial body of legacy finding aids for collections in the Southern Historical


Collection. Although the department has been producing finding aids in EAD since the late 1990s, there has been little systemized retrospective conversion of the legacy finding aids; finding aids have been reprocessed and encoded on an ad hoc basis, depending on transient factors such as staff time, researcher demands on the collection, and grant funding. The professional staff members in the Manuscripts Department were unable to plan for a systematic recon project, however, because they lacked adequate “demographic” information about the legacy finding aid collection. My availability to conduct a complete survey of the legacy finding aids enabled them to begin the process of planning retrospective conversion at the SHC. 29

Survey Design

In order to develop a useful survey instrument, the staff of Public Services and Technical Services convened a meeting with me to discuss the kinds of information that they felt imperative to gather about the legacy finding aids. They agreed that they needed two kinds of information about the finding aids: quantifiable summary information and evaluative information.

Summary Information Needs

The summary information I collected was a count of the number of pages in the legacy finding aid that were not already online, the number of items in the collection, the date/s that the finding aid was created, and whether or not there was microfilm associated with the collection, according to the finding aid. Technical Services staff wanted to know how many pages, other than what was already online (e.g., the abstracts), were in each of the legacy finding aids so that they could determine the amount of information that did not exist in electronic format. The number of items in the collection was an important

29 Interview with Public Services and Technical Services Staff, 8 November 2007.
piece of information to draw into the survey because it could serve as a means of determining whether further description, or even reprocessing, was necessary for any of the legacy finding aids and the collections that they represent. If a collection item count were very high, for example, but the finding aid was only one page, some reprocessing or enhanced description could be called for, depending on the nature of the collection. The purpose of recording the original and revision dates for the finding aids was multifold. The style, content, and extent of finding aids change over time in accordance with the SHC’s evolving descriptive practices and over the course of the twentieth century the SHC had over seven different formats for their finding aids.\textsuperscript{30} The date that a finding aid was created reflects the style in which the finding aid was written, which provides some measure of the breadth and depth of information contained therein. The legacy collection of finding aids contains quite a few “cut and paste” documents; these finding aids are the amalgamation of description written in different periods, therefore have multiple dates that represent their years of creation. Finally, the purpose of recording whether or not the legacy finding aid indicated the presence of microfilm served a control function. For a number of collections the only record of associated microfilm is a green sheet in the legacy finding aid; Public and Technical Services wanted me to note the presence of microfilm as a means of getting that information into the Manuscripts Department’s master database, making it more accessible and searchable.

\textit{Evaluative Information Needs}

The evaluative information that I collected was represented by a set of descriptive code letters that I assigned, as relevant, to each of legacy finding aids. The code structure was developed as a result of a preliminary survey that I conducted of the finding aids

\textsuperscript{30} Laura Knodel, “The Evolution of Archival Description at the Southern Historical Collections”, 16.
contained in the first four of the 137 binders that house the legacy description. A pass-
through of the first four binders revealed that their contents fell fairly easily into a few
descriptive categories. All of the collections in the SHC that are not in XML (EAD),
HTML, or ASCII are already represented online by a summary paragraph. These
summaries can also be found in hardcopy form in almost all of the legacy finding aids, as
they were produced as part of a National Endowment for the Humanities-funded
cataloging and re-housing project undertaken in the early 1990s. All of the online
summary records contain an instructive note to the user that a more detailed finding aid
exists in hardcopy in Manuscripts, and this note is followed by contact information for
Public Services. While all of the summary records indicate that more complete finding
aids exist in hardcopy form, this is not the case for all of the collections—I found in my
preliminary survey that the online summary represented essentially all of the information
contained in the hardcopy finding aid in a good number of cases, making the instructive
note misleading. On the basis of this finding, it was decided that “true” and “false”
descriptive codes were needed to categorize the summary record finding aids: “true”
would be applied to those records for which more information did exist in the hardcopy
finding aid; “false” would be applied to those records for which the hardcopy finding aid
contained no significant information not already represented in the online summary. I
also discovered some hardcopy finding aids that were nothing more than printouts of the
online summary record; these cases were designated by the descriptive code “printout.”

The preliminary survey revealed the need for two further descriptive categories:
“additional info” and a default “what the hell” (named thusly to add some levity to the
project). Since the migration of finding aids online, Technical Services policy has been
to remove the hardcopy finding aid from the legacy finding aid binder when an online version of the finding aid is complete and mounted on the Manuscripts website. This procedure has not always been followed, however, and there are finding aids in the legacy collection that are also partially online. The category “additional info” was developed to describe these cases where there is more information in the hardcopy finding aid than represented in the online finding aid. Because there is no indication online that a more detailed finding aid exists in hardcopy format for these “additional info” finding aids, Public and Technical Services wanted an electronic record of these cases in the Manuscripts database. The final descriptive category, “what the hell” was created as a flag notation: a “what the hell” designation meant that there was something awry with the finding aid, online record, or catalog record for the collection. The most common issue that required this designation was the lack of a MARC record for the collection, an indication that a discrepancy existed between the electronic record keeping and the physical presence of the collection. Like the notation of the existence of a microfilm record in the legacy finding aids, the “what the hell” notation mostly functioned as a means of giving Technical and Public Services more control over the collections.

The descriptive categories were then assigned a single code letter, and I entered the evaluative portion of the survey instrument into a table for my own reference. (See Appendix A.)

**Survey Implementation**

I spent approximately ten hours per week over the course of three months conducting the survey of the legacy finding aids in the Manuscripts Department. There
are 137 binders that house the legacy finding aids; the aids are arranged in alphabetical order by collection creator and range in size from one page to over 500 pages. Technical Services added a query in the Manuscripts Department’s master database (a Microsoft Access database) in which I recorded the data about each legacy finding aid that I surveyed. This table was linked to the master control table so that the name of each collection, its distinct record identification number, and web format of the finding aid were automatically drawn into the survey query. (See Appendix B, Figure 1)

In order to assess the finding aids on the basis of the criteria in the survey instrument, I used the hardcopy finding aids as my starting point. My process for each finding aid was multi-stepped. Working from the legacy finding aids, I looked up the collection number in the OPAC and used the MARC record for each collection to note the number of items in the collection in the database. I then turned back to the hardcopy finding aid and counted pages, recorded dates when they were present and estimated when they were not, and noted if there was a green sheet indicating the presence of microfilm for the collection. In order to apply the one letter descriptive codes, I compared the hardcopy finding aid to its electronic surrogate on Manuscripts’ website to determine whether the summary record adequately captured all the information contained in the finding aid, or if there was more information in the hardcopy finding aid than in the online summary record. For the legacy finding aids that also had finding aids online, I assigned the “additional information” property code if there was description in the legacy

---

31 The control table indicates whether the finding aid exists online in HTML, ASCII, w2 (XML), or summary form.

32 Most of the finding aids had dates associated with them. Because the SHC used approximately seven different styles of finding aid over the course of the 20th century, I was usually able to estimate at least a decade when the finding aid did not have a specific date on it.
finding aid that was of substance and not represented in the online version, or I removed the finding aid from the legacy binder if it was represented in full online. While this weeding function was not an intended purpose of the survey, I was able to remove the hardcopies of 264 finding aids from the legacy binders.\footnote{See the online finding aids at http://www.lib.unc.edu/mss/gm/index.html.}

After I went through the 137 legacy binders and collected the survey information for each finding aid, I was able to run a series of analytical queries on the data. (See Appendix B, Figures 2-4 for examples of queries) I built the analytical queries on the master query that I had entered the survey information into. I also used the collection usage query kept by Public Services staff to produce figures on the usage of the collections represented by the legacy finding aids.

**Survey Results**

My survey revealed that Manuscripts holds 2695 legacy finding aids that need attention from the Technical Services staff. The vast majority of these finding aids, 2459, are online only in summary format, which means they will require Technical Services to convert the hardcopy finding aid into electronic format. The other 136 finding aids that require intervention are in a variety of electronic formats already: ASCII, HTML, and x2 (XML). The majority of the legacy finding aids (2323) are five pages or less in length; 380 are between six and nine pages long, and only 165 are ten pages or longer. (See Appendix C, Table 1)

The majority of the legacy finding aids, 1324, have the true property, meaning that there is more information in the hardcopy description than there is in the online summary record. (See Appendix C, Table 2) Three hundred fifty seven of the legacy finding aids have the false property: the online summary indicates that a more detailed
hardcopy finding aid exists, however this is not true—the hardcopy finding aid has no content other than what is already represented online. Seventy nine finding aids have the additional information property—there is more information in the hardcopy finding aid than represented in the online finding aid, although this is not indicated in the online version. Of the 141 finding aids with a “what the hell” designation, only 13 have an online summary record—the vast majority of the “h” finding aids have no MARC record, meaning they have been incorporated into other collections, de-accessioned, or made otherwise unavailable. One hundred eighty one finding aids have the printout property—the hardcopy finding aid is simply a printout of the online page.

I used Public Service’s collection usage records, also kept in the master Manuscripts database, to generate data about the frequency of usage of the collections represented by the legacy finding aids. (See Appendix C, Table 3) More than fifty percent of the collections represented by the legacy finding aids have been used at least one time in the past five years. The finding aids that have 11+ pages represented the collections most likely to have been used more than ten times over the past five years: 23% of the 11 paged finding aids were used ten or more times, while fewer than 10% of the shorter finding aids were used ten or more times over the same time period.

I also analyzed the legacy finding aid collection by date. Slightly more than half (52%) of the legacy finding aids were produced in one time period: 1386 had only one creation date associated with the archival description. (See Appendix C, Table 4) Most of the finding aids that have only one date associated with them were produced in the early 1990s, under the NEH cataloging grant. There are very few of the original WPA finding aids from the 1930s left in the legacy collection; presumably most of these old
finding aids were re-written in later periods. One thousand two hundred ninety six of the legacy finding aids are comprised of description written in multiple periods; some of the finding aids had as many as five different dates associated with them.

**Discussion of Results**

While the long term goal underlying this survey is the retrospective conversion of each of the legacy finding aids to EAD, I made some suggestions to Technical Services on the basis of the results of the survey and the analysis that I did of the data:

**Recon by Property Characteristic**

It would be possible to begin the migration of the legacy finding aids by eliminating the 79 finding aids with the additional information property; these finding aids are, by in large, online already and only require the addition of some information to be completely online. The 357 finding aids that have the false property could also be a relatively straightforward place to start weeding the binders. The false property indicates that there is no more information in the hardcopy finding aid than in the summary record online; 199 of the finding aids with the false property have one item in the collection and a one-paged finding aid, and Technical Service might safely assume that the current finding aid adequately represents the single item in the collection and quickly convert these finding aids to EAD format. Fifty-five of the false propertyed finding aids have ten or more items in the collection; only eight of these 55 have more than one paged finding aids. Technical Services may need to assess these collections before dealing with the finding aids, as it is possible that some of the collections need further description (e.g., the Oscar Knefler Rice collection, which looks to be unprocessed to date). The 181 finding aids with the printout property could also be easily tackled, as the information
that is online is all that exists, unless Technical Services decided that any of the printout collections need more description. Only 28 of the printout finding aids represent collections with more than ten items contained therein; ten of those 28 are very large, unprocessed collections.

_Recon by Item Number versus Page Length_

Another approach that Technical Services could take would be retrospective conversion based on item number and page length. There are only 189 single-paged finding aids that represent collections that contain more than 25 items. While 25 items may be a somewhat arbitrary value, it may be more important for Technical Services and Public Services to confer about the collections with high item counts and very short finding aids. Some of these collections may require more description in order to make them more interesting and accessible to users. The remaining 1175 single-paged legacy finding aids represent collections with fewer than 25 items and could potentially be converted to EAD with relatively little staff time.

_Recon by Usage_

In addition to analyzing the survey according to page length and descriptive categories, I also did an analysis on the basis of collection usage for the legacy finding aids. The Public Service staff keeps a record of collection usage in the master database, and using this data in conjunction with the data I collected I was able to determine usage of the collections that are represented by the legacy finding aids over the past five years. The purpose of gathering data about the use of the collections that have legacy finding aids was to assist Public Services and Technical Services in making prioritizing decisions about which collections should undergo receive retrospective conversion to EAD first.
The percentage of collections used ten or more times over the past five years is fairly small in each page length category; there are 65 finding aids that are ten pages or less in length for collections that have been used ten or more times over the past five years, and 37 finding aids that are more than ten pages for the same category. In total, this comes to 102 finding aids that, on the basis of usage alone, should be retrospectively converted first. Out of the 2695 finding aids that need conversion, the 102 that have been used more than ten times represents a mere .4%.

**Implementation of Survey Results**

As a result of the survey and the information that it yielded, the staff in the Manuscripts Department have been able to begin planning their retrospective conversion process. They are still in the preliminary planning phases, however, as the survey was just completed at the end of April 2007. Prior to my completion of the survey of finding aids, the staff in Technical Services were aware that they have some 1100 finding aids that are in HTML or ASCII. These finding aids are in standardized enough format to send out for contract processing. They also knew how many finding aids they have that do not have anything other than a summary record representing them in electronic format. They did not, however, know very much about the scope and content of the legacy finding aid collection. With the data from my survey, Technical Services has been able to classify the legacy finding aids—they have descriptive data that is both quantified and evaluative, which gives them the ability to assess the amount of time and resources that will be needed to complete the recon of the finding aids.

Two of the most straightforward pieces of information that I collected—page count and true property (more content in the hardcopy finding than in the online
summary)—have already been put to use in a preliminary application for a digitization grant from North Carolina Exploring Cultural Heritage Online (NC ECHO). The purpose of this grant application is to receive funds to convert all of the legacy finding aids with short page counts and the true property. The staff in Technical Services decided that they could more easily argue for the conversion of quantity over a criteria based on more slippery categories such as usage. Usage, they point out, is a somewhat subjective category, as collections may be used many times by the same user, or overlooked on the basis of insufficient description rather than the quality of the collection. Pagination, while an imperfect measure of finding aid content, provides a neutral accounting of relative size and translates comprehensibly outside the walls of the repository. In the future, they are projecting that they may be able to apply for content-based funding for the longer, more intractable legacy finding aids such as the Arthur Raper Collection finding aid, or Howard Odum Collection finding aid; because the longer legacy finding aids tend to be for more well-know individuals and organizations, the possibility that funding could be secured on the basis of the prominence of the creator of the collection is greater.

To date, Technical Services has also used the survey data in conjunction with other in-house data to compile a list of legacy finding aids that are relevant to African American history and that are simple enough to retrospectively convert if the funds become available from in-house sources. While the decisions about which specific collections would be most desirable to convert were largely influenced by input from

---

34 For information on NC ECHO’s digitization grants, see http://www.ncecho.org/grantinfo.asp.

35 Interview with Public Services and Technical Services Staff, 8 November 2007.
Public Services staff (on the basis of usage and relevance to a set of topics including the Civil War and Reconstruction, slavery, race relations, 19th and 20th century politics, diplomacy and law), the survey data was marshaled to assess the feasibility, particularly in reference to number of pages, of converting the chosen finding aids.36

Conclusions

The survey that I conducted was designed to meet the specific information needs of the staff in the Manuscripts Department for their retrospective conversion planning, but it could be instructive for any institution or repository that is similarly confronted with a large body of non-EAD compliant legacy finding aids. Because wholesale retrospective conversion work is substantively labor and cost intensive, the risks and costs associated with failure preclude off-the-cuff encoding. Moreover, because most repositories have finding aids that were produced before the current incarnation of archival descriptive standards as well as before the arrival of computers (and easily manipulated digital files), proactive surveying of the finding aids and description is certainly in the interest of all involved. The survey could be taken before an EAD implementation occurs, as part of an institution’s larger-scale planning for digitization, or, in the case at the Manuscripts Department at UNC-Chapel Hill, as part of an institution’s efforts to move older archival description online. Surveying the finding aids that need conversion allows for the informed placement of the recon project in the continuum of work that faces all processing archivists. With ever-growing processing backlogs pressing behind the scenes at many archival repositories, and users pressing up front for more detailed and exhaustive web-based information, the busy archivist should

36 Interview with Public Services and Technical Services Staff, 8 November 2007.
spend some quality time quantifying and planning any recon before typing one angley bracket.
### Evaluative Descriptive Codes

<table>
<thead>
<tr>
<th>Code</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>M</strong> (microfilm)</td>
<td>The hardcopy finding aid indicates there is microfilm associated with the collection</td>
</tr>
<tr>
<td><strong>T</strong> (true)</td>
<td>There is more information in the hardcopy finding aid than there is in the online summary finding aid</td>
</tr>
<tr>
<td><strong>F</strong> (false)</td>
<td>There is not more information in the hardcopy finding aid than there is in the online summary finding aid</td>
</tr>
<tr>
<td><strong>H</strong> (&quot;what the hell&quot;)</td>
<td>There’s something amiss about the collection record, the hardcopy finding aid, or the online finding aid</td>
</tr>
<tr>
<td><strong>P</strong> (print out)</td>
<td>The hardcopy finding aid is a print-out of the online finding aid</td>
</tr>
<tr>
<td><strong>A</strong> (additional information)</td>
<td>There is information in the hardcopy finding aid that is not represented in the online finding aid; the online finding aid does not indicate that more information exists elsewhere</td>
</tr>
</tbody>
</table>
APPENDIX B

Figure 1. Screen Shot of Survey Instrument Query and Query Design
APPENDIX B, Continued

Figure 2. Screen Shot of True Property Query and Screen Shot of Query Design
APPENDIX B, Continued

Figure 3. Screen Shot of Page Number Query and Query Design
Figure 4. Screen Shot of 25+ Items versus 1 Paged-Finding Aid Query and Query Design
Survey Results in Tabular Format

**Table 1. Finding Aids by Page Length**

<table>
<thead>
<tr>
<th>Page Length</th>
<th>Number of Hardcopy Legacy Finding Aids</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 page</td>
<td>1364</td>
</tr>
<tr>
<td>2 pages</td>
<td>377</td>
</tr>
<tr>
<td>3 pages</td>
<td>286</td>
</tr>
<tr>
<td>4 pages</td>
<td>165</td>
</tr>
<tr>
<td>5 pages</td>
<td>109</td>
</tr>
<tr>
<td>6 pages</td>
<td>72</td>
</tr>
<tr>
<td>7 pages</td>
<td>42</td>
</tr>
<tr>
<td>8 pages</td>
<td>36</td>
</tr>
<tr>
<td>9 pages</td>
<td>35</td>
</tr>
<tr>
<td>10 pages</td>
<td>13</td>
</tr>
<tr>
<td>11+ pages</td>
<td>141</td>
</tr>
</tbody>
</table>

**Table 2. Finding Aids by Descriptive Code**

<table>
<thead>
<tr>
<th>Descriptive Code</th>
<th>Number of Hardcopy Legacy Finding Aids</th>
</tr>
</thead>
<tbody>
<tr>
<td>True</td>
<td>1324</td>
</tr>
<tr>
<td>False</td>
<td>357</td>
</tr>
<tr>
<td>Additional Information</td>
<td>79</td>
</tr>
<tr>
<td>“What the Hell”</td>
<td>141</td>
</tr>
<tr>
<td>Printout</td>
<td>181</td>
</tr>
</tbody>
</table>

**Table 3. Finding Aids by Collection Usage, 2002-2007**

<table>
<thead>
<tr>
<th>Length of finding aid</th>
<th>1 or more times</th>
<th>5 or more times</th>
<th>10 or more times</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 page</td>
<td>743 (54%)</td>
<td>61 (4%)</td>
<td>11 (0.8%)</td>
</tr>
<tr>
<td>2 pages</td>
<td>261 (69%)</td>
<td>52 (14%)</td>
<td>11 (3%)</td>
</tr>
<tr>
<td>3 pages</td>
<td>212 (74%)</td>
<td>45 (16%)</td>
<td>6 (2%)</td>
</tr>
<tr>
<td>4 pages</td>
<td>118 (72%)</td>
<td>39 (24%)</td>
<td>8 (5%)</td>
</tr>
<tr>
<td>5 pages</td>
<td>89 (82%)</td>
<td>31 (28%)</td>
<td>7 (6%)</td>
</tr>
<tr>
<td>6 pages</td>
<td>73 (83%)</td>
<td>21 (24%)</td>
<td>10 (1%)</td>
</tr>
<tr>
<td>7 pages</td>
<td>42 (84%)</td>
<td>11 (22%)</td>
<td>3 (.6%)</td>
</tr>
<tr>
<td>8 pages</td>
<td>36 (95%)</td>
<td>14 (37%)</td>
<td>2 (.5%)</td>
</tr>
<tr>
<td>9 pages</td>
<td>35 (90%)</td>
<td>14 (36%)</td>
<td>5 (13%)</td>
</tr>
<tr>
<td>10 pages</td>
<td>13 (76%)</td>
<td>5 (29%)</td>
<td>2 (12%)</td>
</tr>
<tr>
<td>11+ pages</td>
<td>141 (88%)</td>
<td>75 (47%)</td>
<td>37 (23%)</td>
</tr>
</tbody>
</table>
APPENDIX C, Continued

Survey Results in Tabular Format

Table 4. Finding Aids by Decade

<table>
<thead>
<tr>
<th>Decade</th>
<th>Number of Hardcopy Legacy Finding Aids Produced Therein</th>
</tr>
</thead>
<tbody>
<tr>
<td>1930s</td>
<td>5</td>
</tr>
<tr>
<td>1940s</td>
<td>49</td>
</tr>
<tr>
<td>1950s</td>
<td>65</td>
</tr>
<tr>
<td>1960s</td>
<td>16</td>
</tr>
<tr>
<td>1970s</td>
<td>12</td>
</tr>
<tr>
<td>1980s</td>
<td>34</td>
</tr>
<tr>
<td>1990s</td>
<td>1023</td>
</tr>
<tr>
<td>2000s</td>
<td>182</td>
</tr>
</tbody>
</table>
BIBLIOGRAPHY


Whitaker, Beth M. “Get it, Catalog It, Promote It: New Challenges to Providing Access,” RMB 7, no. 2 (Fall 2006):121-133.