

2008

Process and Product: Jump-Starting Archival Cataloging

Chatham Ewing
University of Mississippi

Follow this and additional works at: <https://aquila.usm.edu/theprimarysource>

 Part of the [Archival Science Commons](#)

Recommended Citation

Ewing, Chatham (2008) "Process and Product: Jump-Starting Archival Cataloging," *The Primary Source*: Vol. 27 : Iss. 2 , Article 4.

DOI: 10.18785/ps.2702.04

Available at: <https://aquila.usm.edu/theprimarysource/vol27/iss2/4>

This Article is brought to you for free and open access by The Aquila Digital Community. It has been accepted for inclusion in The Primary Source by an authorized editor of The Aquila Digital Community. For more information, please contact Joshua.Cromwell@usm.edu.

Process and Product: Jump-Starting Archival Cataloging

Chatham Ewing, Digital Initiatives Librarian for Special Collections, The University of Mississippi

Two years ago, the University of Mississippi Department of Archives and Special Collections hired an archivist charged with moving the department's finding-aids onto the Internet. At that time, the information about our collections on the Internet was less comprehensive than the print information in the reading room. Available on the departmental web-site there were a few finding-aids that indexed important collections, subject-based lists of our holdings (designed to allow our patrons to discover brief summaries of many of our most important collections), and general departmental information. Additionally, there were several online exhibits.

The department had good information describing most of our holdings. Researchers were able to consult a variety of finding-aids, subject guides, and a subject index to our collections. But the information was in print form and therefore only available to readers who visited the archives in person. Researchers with substantial projects, unless assisted by a local researcher, had to request a print version of our finding-aids before coming to our reading room to work with our collections.

It was clear that our Internet resources as they stood at that time were useful to our patrons, and our patrons indicated that they were pleased by the efforts the department had made in the area of digital information on the Internet; however, our patrons were also making us aware that they wanted to have even more and better descriptions and even more and better digital objects available online. Of course, our patrons, driven by their passion for the research areas our archive supports, are no more demanding than patrons anywhere else. Their requests were not unreasonable given their experiences at other archives.

Our patron's demands encouraged us to consider what to do - and the department discussed this at several staff meetings. After our initial internal discussions, it was clear that the department had three essential concerns: We wanted to ensure that we disseminated good quality collection level information, that our good quality information was found and used by our patrons, and that we accomplished goals one and two in a timely fashion.

The department developed a plan where we took stock of where we were, defined our concerns, considered approaches adopted by other institutions, surveyed the literature written about these issues, developed a strategy for moving forward, and implemented the strategy. What follows will be an outline and discussion of the above six steps taken to reach our goal of good information widely and swiftly disseminated. Though at times we felt as if we were trudging up the Ogre's mountain in the old Irish fairy tale, taking two steps forward and three steps back, after a few false starts and missteps, we feel we have managed to improve our patron's chance of locating information about our holdings, and we feel that once they've located the information it will be good quality information.

Taking Stock and Defining Our Concerns

When we took stock, we found that we were both further ahead and further behind with our collection level information than we thought we were. We had many thorough print finding-aids for our materials, and in many cases these same finding-aids were already in digital form as Microsoft

Word documents. These finding-aids were often quite detailed, even including item level descriptions, particularly in the case of heavily researched and important collections. This was good. We were hopeful that we might be able to easily move our finding-aids from MSWord to a more acceptable form of digital document for online presentation, in spite of the potential pitfalls.¹

However, though we did have digital versions of our finding-aids, and most of these finding-aids served admirably as internal and reading-room reference tools, when we reviewed them we had concerns about presenting them directly on the Internet. As many in the archival community know, legacy finding-aids tend to vary greatly in form and presentation - ours were no exception. This is a problem endemic to converting legacy finding-aids. It certainly is not sensible to expect that finding-aids produced before the American archival community had settled on either a data-content or a data-transfer standard for archival description would conform to any standards.

Finding-aids created before Henson's 1989 Archives Personal Papers, and Manuscripts (APPM) might have clear descriptions that nevertheless appear today to be significantly non-standard in content. Further, APPM was a standard designed with the old MARC-AMC record type in mind (a record type eliminated by format integration). Luckily, before 2005 the University of Mississippi (along with most archives in the United States) did not create online records for its archival collections using MARC records - so we would have no concerns about legacy MARC records. It is also true that while the first edition of International Standard Archival Description, Revision G (ISAD-G) came out in 1994, it was not adopted here at the University (though that was the case with most U.S. archives at that time). Finally, the first edition of Describing Archives: A Content Standard (DACS), the first broadly adopted content standard for description in the U.S., came out quite recently, in 2004. Frankly, it is only within the last few years that U.S. archivists have settled on what they believe a finding-aid's content should look like.

Additionally, when we started our project in 2005, EAD was still only six years old and had been only recently established as a transfer standard for finding-aids. Where MARC had been around for many decades as an acceptable record structure for managing and transferring bibliographical information about books from catalog to catalog, it was not until the 1999 publication of Encoded Archival Description Application Guidelines : Version 1.0, that a widely available standard record structure for finding-aid documents used in data exchange became available. Finally, at that time (and even now) there is no complete standard turn-key system offered as a product by any vendor for loading and managing these kinds of EAD records.

In short, because the standards have developed so recently, the majority of the documents describing our collections understandably showed a disparity from current standards for content and data-transfer structure. We had the happy opportunity to implement the new content standard on our post-2005 finding-aids, develop a plan for what to do with our legacy finding-aids, and implement our own system for delivery.

If our descriptions were to be shared with other institutions and systems, we needed to address this disparity between some of our legacy information and current professional practices with regard to data creation, management, and sharing. Would we attempt to revise the good older information and normalize it with regard to current standards? Would we revise all of the old guides completely, even to the item level? Of course, we also wanted to get information about our collections, even if it only offered the barest minimum of data elements, into the hands of our patrons in a timely fashion

Additionally, while there was a significant subset of more recent finding-aids in which we felt the information was good and in which we only would have to address some questions of data content and data structure, there was another subset of collection descriptions, including mostly those that

were some years old, that were going to need revision in order to synchronize the current state of the collection with the current state of the description. For some of the collections there had been further accessions, for some there had been some rearrangement in the past that resulted in multiple descriptions, and for some there were simply disparities between the arrangement of the physical objects and the description that needed resolution. Updating descriptions meant that we would be doing systematic sampling and retrospective conversion on those collections that failed our sampling test and therefore had finding-aids about which we had concerns. That was going to take staff-time and other resources, and there was no way to avoid doing it.

Surveying the Literature

Thankfully, relevant developments had occurred in the archival world just previous to our beginning this project, and these developments contributed to the department's thinking on these issues. First, the Association of Research Libraries (ARL) had its "Exposing Hidden Collections" conference in 2003, the report of which indicated that directors should seek to have their Special Collections and Archives get a basic collection level record for every collection, even minimally processed collections, available on the Internet.² There was the argument that less robust records might result in greater access for more collections, and the leading voices in this argument were Dennis Meissner and Mark A. Greene.³

At that time, a survey of the state of research in archival informatics included some significant focus on how patrons of archives discover information about our holdings. There had been research on the impact of search engines on discovery done by Helen Tibbo.⁴ In addition, Daniel Pitti and Wendy Duff had edited a compilation of work on EAD.⁵

When we correlated the minimal information we found about archives-specific research with then-contemporary research on the information seeking patterns adopted by general library patrons, certain important axioms for our project became clear. Research indicated that:

- * Patrons tend to use Internet search engines first, just as they begin to search for information.
- * While they may begin with search engines, patrons have a variety of additional information seeking patterns. Since they will seek for information in a variety of different ways, we will have to consider how to deliver it in a variety of ways.
- * Advanced users of archives tended to be more senior scholars and still tended to make extensive use of library catalogs, particularly the large union catalogs. Either the patrons used these tools themselves or archivists used these tools to help patrons.

Other Institutional Practices

Our review of other institution's practices with regard to MARC cataloging and EAD descriptions included a look at the New York Public Library, Washington University in St. Louis, and The University of Illinois. At the time, NYPL was creating collection level records in RLIN and its own OPAC and delivering EAD finding-aids using a Dynatext/Dynaweb. NYPL was using a very powerful system for managing SGML and XML that was no longer commercially available and would not be affordable for us if it were. The University of Illinois used a database backend to assist with the delivery collection level information linked to EAD finding-aids. Washington University in St. Louis, like many universities then and now, was using a strategy that involved cataloging

collections on OCLC and creating EAD finding-aids for web delivery off of the library web-site and through what was to become ArchivesGrid.

Developing a Strategy

While several other possibilities have opened up in the interim, our analysis at the time of potential avenues for discovery of our collections included the following:

1. Internet Search Engines and Indexes (Google, DMOZ, etc.)
2. The University of Mississippi OPAC
3. Regional Catalogs (Kudzu)
4. National and International Catalogs (OCLC's WorldCat, RLIN, NUCMC)
5. Finding-Aid Aggregations (ArchiveGrid, ArchivesUSA,)
6. Authoritative Web-Sites (VoS, Academic Subject Guides, etc.)

A strategy that would blanket as many possible resources and points of access would make our collections most likely to be discovered by potential patrons, and hence the broad nature of potential resources for discovery. While the number of options initially seemed unworkable, some further thinking about this clarified and simplified our approach.

On our list above, we had several MARC based information resources (2, 3, & 4). As with most institutions, our database management department created bibliographic records for print materials on OCLC and downloaded them into our local OPAC. If we could find a way to duplicate this process for records about our archival collections, this would mean that our local OPAC and OCLC could be considered vehicles for essentially the same information. Our regional catalog aggregated information from our and other member's local OPACS using Z39.50. Again, we reasoned that if we could somehow get archival collection level records into OCLC, this MARC -based information would also be re-used through Z39.50 in our regional catalog.

On our list above we also had resources amenable for presentation through using EAD (1, 5, 6). When we posted our EAD based finding-aids on our web-site, search engines such as Google and Yahoo! would discover them. When authoritative resources such as Archives USA, open indexes such as DMOZ, or other online resources like Wikipedia pointed at our web-site, they would be indexing an EAD-based document for the information seeker.

After analyzing how we thought our patrons might be most likely to discover information about our collections, the department came to the conclusion that we initially needed to produce two kinds of information to create the best chance of a patron discovering our collections - an EAD finding-aid and a MARC record on OCLC and/or RLIN. If we could successfully convert from one record format to another, we might only need to create one kind of information. After this first step, we would have to comb through various general purpose indices (such as DMOZ), subject specific indices (such as the online Congressional directory), and general reference sources (Wikipedia) in order to create links to our materials.

Implementing our Strategy

We felt that our initial step in our plan should be to improve the number and richness of our finding-aids on the Internet. This presented us with some problems - with regard to delivering finding-aids to our patrons, we didn't have the funding to obtain the programmers to implement the kind of database back-ends they had at NYPL and at Illinois; however, even with the potential pitfalls of open source software, we might have considered that route now if either Archon and Archivist's toolkit had been more robustly developed than they were then. Given that, we thought at the time

that it would be easiest to present finding-aids online using HTML generated by modified XSLT scripts derived from Michael Fox's EAD Cookbook.

We ran some initial experiments using III's ("Triple I") Meta-Data Builder, but we found that the hoped for accessibility offered by incorporating the finding-aid as a digital objects into the III package didn't pan out. Our finding-aids took an inordinate amount of time to load through the III product, and this caused some consternation amongst our staff. We were unsure as to whether the slow loads of the finding-aids were due to the nature of the database delivering the finding-aid or due to the large amount of conditional logic in our stylesheet. Eventually we punted, declared it a little bit of both, and, not without a bit of sadness at letting such a promising tool go by the wayside, decided not to use the III product as a delivery mechanism. Instead, we would pre-process our finding-aids into static XHTML files and place them in a directory on our web-server.

This choice to use XHTML was also a result of our initial literature survey. The literature made it obvious that an initial goal should be getting collection-level records describing our holdings up on the Internet where they would be exposed to Internet search-engines. A subsequent goal would be getting collection level records into OCLC and RLIN. A limiting factor for this plan was the state of the descriptions of our collections - it was clear that we had two kinds of collection descriptions.

The first were complete, accurate, and modern - this group consisted of about 70 collections and was augmented through descriptions prepared with the assistance of the University of Southern Mississippi's statewide Civil Rights project. This first group would be the easiest to pull into EAD finding-aids and collection level records. These were the collections that we would use to initially establish our footprint on the Internet (and subsequently on RLIN and WorldCat).

The second included descriptions that were older, many of which were high-level box and folder listings lacking all but the most basic information about collections. This is the group that was going to take some research and retrospective conversion. We simply did not have the resources within the department to do the labor intensive work to produce detailed descriptions of every level of every collection that didn't have it already. The department came to the conclusion that a feasible process would produce "stub" records of all collections using minimal EAD finding-aids, place these records in a searchable directory on our web-site, and pursue collection-level records in WorldCat and RLIN for large, heavily used, and well-described collections.

This would be a good initial step toward meeting the Hidden Collections guidelines. Our initial "stub" records would contain minimal collection level information, including the following information:

- * Title
- * Language of Materials
- * Extent of Materials
- * Unique Identifying Number
- * Physical Location
- * Repository Information
- * Citation Information
- * Use Restrictions.

As soon as we could we would add:

- * Creators
- * Inclusive Dates

- * Abstract and/or Scope Notes
- * Biographical/Historical Notes.

While these elements are far from a complete archival description, they give enough of the basics about a collection to allow a researcher to assess whether the collection might meet her research needs.

Our first step in this process included working with the Mississippi Digital Libraries Civil Rights project to convert some of our existing finding-aids from Word documents to EAD finding-aids, and beginning to present some EAD tagged finding-aids on our own web-site as HTML files generated through the use of XSLT. Needless to say, this amount of conversion and retrospective conversion required and continues to require a good deal of effort and expertise, and our staff has been heavily committed in the area of retrospective conversion.

As mentioned above, we had initially hoped to use our stub EAD finding-aids to generate collection level information to be placed into the largest of the bibliographic databases (RLIN and WorldCat). Our overly optimistic thinking on how to do this was inspired by a remark by Michael Fox, made at an SAA meeting several years ago, about using MARC records as the foundation for all the EAD finding-aids at the Minnesota Historical Society. Since we had finding-aids and no MARC records, what would happen if we could go the other way - using a finding-aid as the basis for a MARC record? With Terry Reese's MarcEdit in mind, we had the initial notion that we were going to pull collection-level records from our finding-aids into MARC, and then upload those as a batch into WorldCat. While this process was technically feasible, and potentially efficient, we were unable to resolve our workflow for doing this plan in workable fashion, and so it was abandoned.

The difficulty we had with converting our finding-aids directly into MARC, and our lack of experience with collection level MARC records, eventually led us to conclude that we might better seek expertise and training for creating proper collection level MARC records outside of our library. The first place that we contacted after coming to this conclusion was the Library of Congress' National Union Catalog of Manuscript Materials (NUCMC) office.

The NUCMC program began publishing a comprehensive index to archival and manuscript holdings in the United States in 1959. This published version covered archival collections registered through NUCMC dating from 1959-1993. Numerous library subject pages describe the resource as a tool for accessing information about national archival collections. In 1986, well after the first wave of library automation discussed by Kilgour⁶, the NUCMC office began a program to help catalog the holdings of repositories that did not have the resources to get collection-level records cataloged on RLIN and WorldCat.⁶ By 1993 the print version of the catalog was no longer produced, and archivists, somewhat behind the general library digitization curve, had moved the NUCMC program entirely into the digital realm. Harriet Ostroff and Claire Gabriel discuss this history.^{7,8}

At the moment the NUCMC catalog operates much as it has for many years - it is an Internet gateway that uses a simple search interface on aggregated collection level records prepared by the program. It is a window into a subset of archival records now held in WorldCat. Somewhat more recently, Chaydwyk-Healey has incorporated the entries in the old print NUCMC into its ArchivesUSA digital product - retrospectively digitizing the old volumes and creating a powerful new tool for access. Of course their program, which also includes NIDS records, is a commercial product which charges subscription fees.

The administrators of the current NUCMC program found merit in our appeal for assistance. They

agreed with our proposition that because our archive did not have staff experienced in the preparation of collection-level records, and our general library cataloging department was overburdened with work, we would need the help of the NUCMC program in order to create proper access to our collections.

After NUCMC gave us the green light, we submitted information about our collections (drawn from our finding-aids, administrative files, and the collections themselves) into their online forms on the NUCMC site at the Library of Congress. They then used our information to create MARC records. After the records were created, NUCMC sent us the records for review. At times there was a correction, though not often, and soon after our approval, the record was finalized for the NUCMC catalog.

As a result, during the last several months we at the University of Mississippi have used those online forms to participate in the NUCMC program. During the course of several months we have been part of the program, the pace has averaged out to about a record a week. Different factors influence how long producing a particular record might take. At times records were easily produced because we had complete information, and at times we needed to do some information gathering within the collections and within information resources in order to assist the NUCMC catalogers. We have communicated regularly with the catalogers at NUCMC over the phone and through e-mail.

Our interaction with NUCMC has proved to be tremendously helpful for us. Because of NUCMC, we have been able to get high quality collection-level records created and have added collections to the largest of the library union catalogs, RLIN and WorldCat (though now only WorldCat, as OCLC has finally swallowed RLG and the two catalogs have been integrated). This has helped us further our strategy of getting high-quality information about our archival and manuscript collections pushed out in a variety of media. Further, NUCMC has often provided additional resources for doing authority and subject cataloging that we have then been able to re-introduce into finding-aids.

We currently have plans to work with NUCMC on the cataloging of many more of our collections, and hope to be presenting better and more information about our collections on WorldCat, our web-site, and on linked web-sites in the future.

Assessment and Conclusions

The recent data-conversion work at the University of Mississippi has greatly increased the amount of information about collections now available both on the Internet and in library databases. We have already had many patrons discover our collections through our newer Internet finding-aids, and our hope is that continuing to post stub and full finding-aids and participation in NUCMC will continue to improve our patron's experience of our collections. They get better information and more of it. Not only is finding-aid use easier to track, but reference work (both in our reading room and with correspondents) can be more efficient and exact.

While the conversion and presentation of our existing information may take us another year or so, some subsequent projects for this data (and our digital collections in general) will involve developing answers to the following questions:

- * How our collections data can be used for cross-linking resources on the Internet?
- * Which authoritative resources we should link to, and how should we manage those links?

- * Will something like a URN be necessary for our finding-aids?
 - * How should we integrate/link our finding-aids with our online digital projects?
 - * How we can develop more efficient processes for creating our metadata?
 - * How we can make our online resources more interactive and patron friendly?
-

Endnotes

¹ Meissner, Dennis. "First Things First: Re-engineering Finding Aids for Implementation of EAD," *American Archivist* 60.4 (1997).

[return to text]

² Exposing Hidden Collections: 2003 Conference

Summary http://www.arl.org/rtl/speccoll/hidden/EHC_conference_summary.shtml

[return to text]

³ Greene, Mark A. and Dennis Meissner. "More Product, Less Process: Revamping Traditional Archival Processing," *American Archivist* 68.2 (2005).

[return to text]

⁴ Tibbo, H. R., et. al. "Finding finding aids on the World Wide Web," *The American Archivist* v. 64 no. 1 (Spring/Summer 2001) p. 61-77.

[return to text]

⁵ Pitti, Daniel and Duff, Wendy. *Encoded Archival Description on the Internet*. Haworth Information Press, 2001.

[return to text]

⁶ Kilgour, Frederick G. "History of Library Computerization," *Journal of Library Automation* 3.3 (1970).

[return to text]

⁷ Ostroff, H.. "Subject access to archival and manuscript materials." *American Archivist*. 53. (1990) p. 100-105.

[return to text]

⁸ Gabriel, Claire. "Subject Access to Archives and Manuscript Collections: An Historical Overview," *Journal of Archival Organization*, Vol. 1 Issue 4 (2002) p 53-63.

[return to text]

Chatham Ewing works in the Department of Archives and Special Collections at the University of Mississippi as the Digital Initiatives Librarian. He currently serves on the Board of the Society of Mississippi Archivists and is a member of the DACS working group of the Society of American Archivists.
