

---

## Creating an Archives Management System at the University of Maryland Libraries

---

**Jennie A. Levine** (*levjen@umd.edu*)

*University of Maryland*

**Amit Kumar** (*kumaramit01@gmail.com*)

*University of Illinois at Urbana-Champaign*

**Susan Schreibman** (*sschreib@umd.edu*)

*University of Maryland*

**Jennifer Evans** (*jennifer.evans@nara.gov*)

*National Archives and Records Administration*

---

**E**ncoded Archival Description (EAD) is an XML-based standard used to encode archival finding aids that reflects the hierarchical nature of archival collections and that provides a structure for describing the whole of a collection, as well as its components (Pearce-Moses 2004).

Archives approach EAD in different ways. The tools available through the official EAD website at the Library of Congress (<http://www.loc.gov/ead/>), such as the EAD Cookbook, focus on making it possible for archivists to tag their existing paper/word-processed finding aid content. This approach works well for smaller institutions, but with over 400 finding aids, manual tagging was not a viable solution at the University of Maryland.

In order to help define the ideal system for our institution, the staff of the Archives and Manuscripts Department identified several main themes. Streamlining workflow was the largest focus. Finding aids are the natural end result of a series of archival procedures that begin with appraisal, selection, arrangement, and finally, description. Since these processes are related, the creation of a system to help the department streamline workflow while also simplifying and demystifying the creation of an EAD document seemed the most practical course of action. Standardization was another concern. We wanted to create a system that organized our finding aid data in such a way that a future generation could easily port it to a new system. Search capability was a third requirement. Searching across collections would greatly aid us both with our reference services and with our processing. Fourthly, we wanted to create a distinct, usable, and attractive public interface.

The department had already created a database in Microsoft Access to keep track of basic information, such as collection title, collection size, and location. As a first step, we modified this database to hold several more collection-related fields and also added a report that allowed staff to create electronic accession sheets (the first step in the record-keeping process for any collection acquired by the department).

Since much of the accession sheet information becomes part of the later EAD document, the new database tables and structure were based on the structure of EAD. Many of the fields were named after their corresponding EAD tags.

The database structure is relatively simple. A main table, named *archdescid* after one of the main components of an EAD document, contains the bulk of the information. A handful of smaller tables tie together the `<eadheader>` information and the deeper-level descriptive information located in the `<dsc>` sections of the EAD document. The biggest challenge was figuring out a way to design a table that would accurately represent the "Box Inventory" section of the paper finding aids so that data entry was simple for staff, but that would also easily convert into the EAD tag structure.

The decision to use *Microsoft Access* as the primary database was based on a number of rationales, although the primary ones were staff familiarity and widespread availability of the product within the institution. Several other institutions use web forms for entering EAD information into a database, and while this method is very flexible and allows the system to be easily shared across institutions, it would not allow the department to carry out some of the other collection management tasks.<sup>1</sup>

In the absence of a skilled programmer, *Microsoft Access* adequately served the purposes of the early phases of the project. Little to no programming expertise was necessary to create functional database forms and reports. There were, however, some weaknesses in the Microsoft Access software that put the project on hold at a crucial point: the plan to create the EAD document using the Microsoft Access report features would not work; the reports could not handle the text in large memo fields. A model for the conversion from Microsoft Access to EAD came from the Australian Heritage Document Management System, which was created by the Australian Science and Technology Heritage Center. It also used an advanced Microsoft Access database and ASTHC staff was helpful in discussing the system. After examining their approach, the University of Maryland realized that the assistance of a programmer would be needed to properly extract the data from Microsoft Access.

The Archives and Manuscripts Department thus approached Maryland Institute for Technology in the Humanities (MITH) to assist with the programming support needed, as well as the project management skills to convert the Microsoft Access database into a series of outputs (primarily finding aids and

subject guides), as well as create an online publishing system with a robust search and browse interfaces, and an administrative management system.

The software itself is comprised of two independent systems: a converter program written in Java that communicates with the Microsoft Access database using Java Database Connectivity (JDBC), and a web application with an XML Content Management System.<sup>2</sup> The web application is based on Java Servlet API with Model View Controller architecture.

The converter application creates a list of finding aids in the database and a user can click and generate the EAD-compliant XML document.

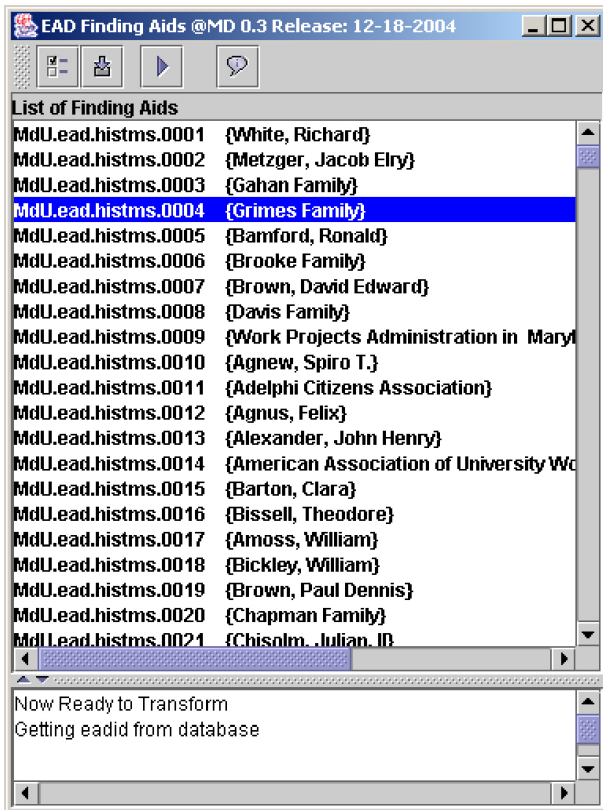


Figure 1: The converter application which transforms the finding aids from the Microsoft Access database into EAD-compliant XML.

These documents are then uploaded and indexed by the web application. The web application also generates the subject guides and finding aids using XSL style sheets.

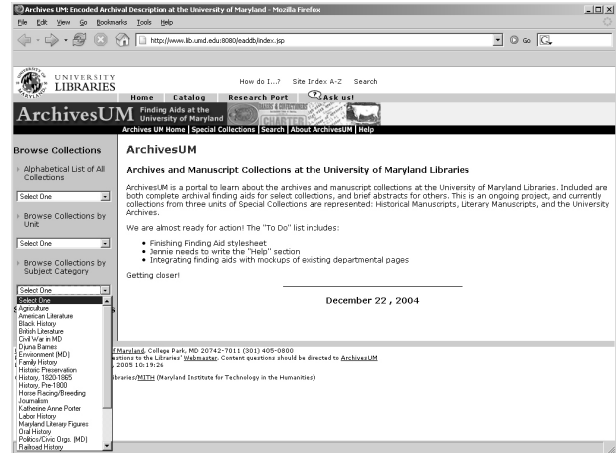


Figure 2: The home page of ArchivesUM, with a pull-down menu listing the subject guides which are generated through a combination of static HTML and dynamically-generated content from the database.

Via the administrative interface, the repository editor can upload, delete, and convert finding aids to HTML. This pre-processing of the XML document was built into the system so that the finding aids did not have to be converted to HTML at the time of the request. Figure 3 shows a result page ranked in order of relevance. It was decided that in the first instance, all collections would be represented in the database through an abstract. As finding aids are converted, they will be made available through the archive management system. As Figure 3 shows, the interface makes it clear when the finding aid is available:

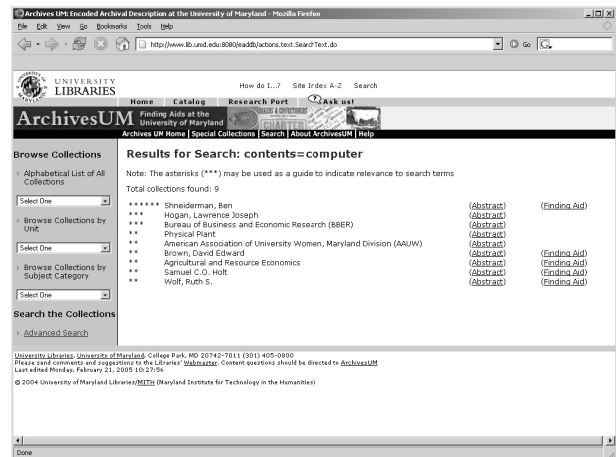


Figure 3: A ranked result page indicating which items have finding aids available.

Generating subject guides proved a greater challenge. Although it would have been easy to generate the subject guides on the fly, it was felt that these needed to be converted into static HTML pages and mounted on the Internet. Subject guides indexed by Google and other search engines has proved to be the most popular way for potential users to find the University of Maryland's archival resources. Thus, a feature was built into the administrative interface to create the subject guides through a combination of static text and abstracts generated from the

EAD document, where <abstract> tags with different "type" attributes are located.

Various nodes of the <eadheader> and <archdesc> are indexed with Lucene and a query interface is provided to search and browse the finding aid.<sup>3</sup> The use of Lucene as a search index enables compound searches for phrases in the box inventory, collection title, author, scope, and subject fields of the EAD document.

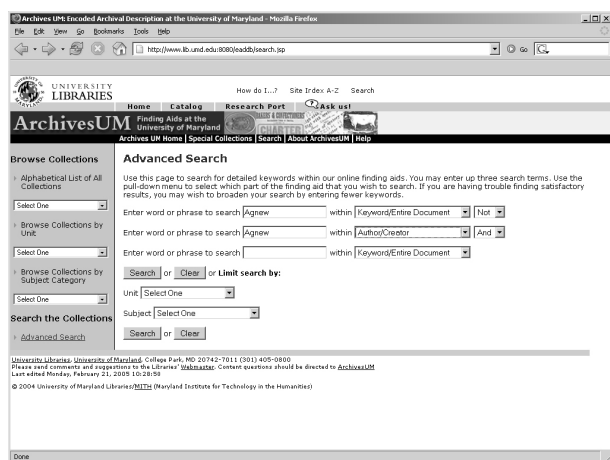


Figure 4: A search page enabling users to perform complex searches based on information in different parts of the EAD finding aid.

The Archives and Manuscripts Department staff and MITH worked together in the development of several XSLT style sheets for various parts of the website. In many ways, this proved to be the most difficult task. The hierarchical nature of the display of a finding aid made design of the final, and most important, style sheet extremely complicated. Other repositories provided examples to build from, but since the EAD of the <desc> section of a document varies widely from institution to institution, advanced customization was necessary.

The administrative interface provides an interface to upload an XSL style sheet, so that the website administrator can change the design of the finding aids and subject guides. Much of the software code for this project has been borrowed from *teiPublisher*. Moreover, although staff developed the system for use with finding aids in three of the units within the University of Maryland's Archives and Manuscripts Department, staff constructed it with the possibility that other archival units on campus could use it, as well as staff in repositories across the University of Maryland system. While each repository will have its own Microsoft Access database so it may generate reports unique to its holdings, there will be one EAD repository, which will give users unprecedented access to search across archival units and institutions in a way not possible currently.

This paper will thus address the theoretical, practical, and programming decisions that contributed to the design of this archival management system.

1. *Virginia Heritage Guides to Manuscript and Archival Collections in Virginia; Online Archive of California.*
2. JDBC <<http://java.sun.com/products/jdbc/>>.
3. Lucene <<http://jakarta.apache.org/lucene/docs/index.html>>.

## Bibliography

Dooley, Jackie M., ed. *Encoded Archival Description: Context, Theory, and Case Studies*. Chicago: Society of American Archivists, 1998.

*EAD Help Pages - Software Products*. EAD Roundtable of the Society of American Archivists, 2003. Accessed 2003-08-13. <<http://jefferson.village.virginia.edu/ead/products.html>>

*Encoded Archival Description (EAD): Official EAD Version 2002 Website*. Library of Congress, 2002. Accessed 2005-03-21. <<http://www.loc.gov/ead/>>

Feeney, Kathleen. "Retrieval of Archival Finding Aids Using World-Wide-Web Search Engines." *American Archivist* 62.2 (Fall 1999): 206-228.

*Heritage Document Management System*. Australian Science and Technology Heritage Center, 2003. Accessed 2003-04-15. <<http://www.austehc.unimelb.edu.au/HDMS/findingaids.html>>

Miller, Fredric, ed. *Arranging and Describing Archives and Manuscripts*. Chicago: Society of American Archivists, 1990.

*Online Archive of California*. University of California, 2004. Accessed 2005-03-21. <<http://www.cdlib.org/inside/projects/oac/toolkit/>>

Pearce-Moses, Richard. "Encoded Archival Description." *A Glossary of Archival and Records Terminology*. Website: Society of American Archivists, 2004. <<http://www.archivists.org/glossary/>>

*Virginia Heritage Guides to Manuscript and Archival Collections in Virginia*. University of Virginia, 2004. Accessed 2004-11-04. <<http://www.lib.virginia.edu/vhp/index.html>>

*teiPublisher*. Accessed 2005-05-19. <<http://teipublisher.sourceforge.net/docs/index.php>>