MODULE 3

DESIGNING DESCRIPTIVE AND ACCESS SYSTEMS

Daniel A. Santamaria

Appendix B: Case Studies

Princeton University Archives

by Daniel A. Santamaria

Seeley G. Mudd Manuscript Library at Princeton University is an example of a medium- to large-sized archival repository that uses a wide variety of tools in order to implement an archival access system. Though tools and procedures are evolving frequently, the access system has been quite successful, delivering descriptive data and digital objects to patrons despite some staffing and resource limitations.

Background

The Seeley G. Mudd Manuscript Library, a division of the Department of Rare Books and Special Collections of the Princeton University Library, houses the Princeton University Archives and a highly regarded collection of twentieth-century public policy papers. The more than 500 collections and 30,000 linear feet of archival and manuscript material in its possession are widely used by local, national, and international researchers. More than 2,000 visitors use the Mudd Library's reading room each year, and its staff field another 2,000 electronic, mail, and telephone inquiries annually.

In Fall 2007, staff at the Mudd Library reached a goal of providing online access to all of the Mudd Library's collections. This initiative involved a number of discrete projects, including several ambitious processing projects and a data conversion project resulting in the conversion of all legacy electronic finding aids to Encoded Archival Description.

Most significantly, basic descriptive data and location and holding information was created for all 335 collections, totaling more than 13,000 linear feet, held within the Princeton University Archives. This data allowed for the completion of collection-level MARC cataloging for all collections lacking descriptive records. The MARC records were then converted to EAD, primarily through the use of XSLT stylesheets and Terry Reese's MarcEdit software. Previously, more than two-thirds of the collections within the University Archives were not represented by any descriptive record online.

With the new EAD finding aids, descriptive records for all of Mudd's collections were discoverable in the Princeton University Library's OPAC, the Department of Rare Books and Special Collections' EAD website, union catalogs and databases such as OCLC's WorldCat and ArchiveGrid, and via common Internet search engines such as Google and Yahoo. As of June 2012, 504 records for Mudd Library collections were available.

Staff continue to add to the descriptive records through the creation of series-, box-, or file-level inventories, and as of spring 2012 all collections larger than 1 linear foot were represented with box- or folder-level inventories. The Mudd Library has also revised accessioning procedures to ensure that both collection-level MARC records and EAD finding aids are produced at the time of accessioning. Our commitment to descriptive standards and willingness to embrace new methodologies were essential in the success of the projects. We view these initiatives as quick, relatively low-cost and non-staff-intensive ways to enhance access to our collections. They have also led us to develop a view of description as an iterative process. We are able to expand or revise descriptions as collections are processed. We also view the descriptive records produced during this process as forming the initial descriptive infrastructure for digital library projects. We plan to use the EAD records to provide access to digital surrogates of material in our collections and to explore additional ways for users to interact with finding aids and the material that they represent.

The primary drawback of these methods is the large number of tools currently needed to accession, describe, and deliver descriptive records and digital content and to generally maintain our baseline levels of service. We continue to seek ways to refine our processes and our descriptive and access systems. In 2012, building on our previous work, we will redesign our EAD finding aid delivery platform, implement Ex Libris's Primo as a discovery layer for special collections material, and implement Atlas Systems' Aeon as a circulation and use tracking tool. These three tools will all leverage the structured descriptive data that we have previously created. We are also closely monitoring the development of ArchivesSpace in the hopes that it will help us further streamline and consolidate our descriptive processes and tools.

Description and Access Activities

Pre-custodial/Pre-accessioning

University departments and outside donors are required to complete a transfer form, which is a Microsoft Excel spreadsheet. In addition to contact information about the donors, we ask that at minimum a title, date, and box number are created for each folder to be transferred to the archives. We also ask donors to indicate if they are aware of any records that contain information that may be governed by federal privacy laws (typically, FERPA in the case of University records) or other sensitive information. The current version of the transfer form can be found online.[61]

Instructions for completing the forms and for contacting the archives are available on the Mudd Library website. We have also been experimenting with a webform created through Google Forms, which asks donors to provide additional contextual or domain-specific information about the material to be transferred. The current version of this form can also be found online.[62]

Tools Used: spreadsheets for descriptive data; Google Forms

Accessioning

Once the library formally takes custody of the material, whether it is analog or digital, an accession record is created in the Archivists' Toolkit. Basic descriptive and administrative data is entered for each accession according to documentation maintained at the library.

While electronic records and digital material are often transferred on physical media such as external hard drives or flash drives, the library has experimented with using document management and file sharing tools in order to transfer custody of records. The University's Webspace service, built on a document management system called Zythos, allows the library to set up a "dropbox" that university offices and departments can use to deposit electronic files.

Description is considered part of accessioning work at the Mudd Library. Each newly accessioned collection is described with a a DACS single-level optimum descriptive record in both MARC and EAD and published in the library's OPAC and finding aids website. Tools for

---

61 http://www.princeton.edu/~mudd/news/transf_donations/transf_instr/transmit.xls
62 https://docs.google.com/spreadsheet/viewform?hl=en_US&formkey=dENydVlHeG9nan
    J6bWpHQ0JyR0MxZ3c6MQ#gid=0

describing newly accessioned collections and collections already at the library are the same and are described below.

Tools Used: Archivists' Toolkit Accessions module; Voyager Integrated Library System; spreadsheets containing descriptive data that are submitted by donors; Webspace/Zythos for electronic file transfer

Description

Creating New Data. Collection-level records are created in both MARC and EAD for all collections held at the library. Generally, collection-level MARC records are created and published in Voyager and then converted to EAD using MarcEdit. Inventories are created for any collection larger than 1 linear foot. These inventories are typically created using the Archivists' Toolkit Resource module unless data already exists in another structured form, such as a spreadsheet created by a donor.

Though this workflow uses a variety of tools, it is currently easier for us to initially create data using our library system and then transform it to EAD, rather than edit MARC-XML records produced by the Archivists' Toolkit. For small edits and additions to finding aids, it is also much more efficient to edit XML files than export resource records from the Archivists' Toolkit. We anticipate streamlining this workflow and moving entirely to the Archivists' Toolkit for description in the next year, primarily because we will index EAD data in Primo, the Ex Libris company's discovery layer product, which will make MARC records in the library OPAC redundant.

Legacy Data. As mentioned above, the library undertook an ambitious data conversion project in 2006 and 2007, resulting in the conversion of nearly 1,000 finding aids from Microsoft Word, WordPerfect, and HTML to EAD. The overhead on the project was substantial; it required writing an RFP and evaluating several proposals, hiring a project manager, and performing quality control on the vendor's work. Overall, however, the project resulted in more finding aids encoded in EAD than could have been achieved through in-house means alone.

The retro-conversion project included only finding aids that existed in electronic form. Mudd Library finding aids that existed only in paper form were encoded using other methods. First, any

collections that lacked collection-level records received DACS's single-level optimum records, using the methods described under backlog reduction projects below.

The records were first created in the library's integrated library system and then exported and converted to EAD using MarcEdit. Collections that were described in paper and typescript inventories were rekeyed, with data entered directly into the Archivists' Toolkit. Though we initially believed this option would not be particularly efficient, we found that the inventories that had never been converted to electronic form were wildly inconsistent and not standards-compliant. Performing the data entry for these problematic descriptions in house allowed archivists to evaluate and restructure the information on an ongoing basis. Combined with the fast rates of data entry achieved by the undergraduate students doing the work, the rekeying was a useful option for us.

Tools Used: Archivists' Toolkit Resource module; Voyager Integrated Library System; oXygen XML Editor; MarcEdit; spreadsheets for inventory conversion for specific projects (often large or complex); Syncro SVN client

Backlog Reduction and Collection Management Projects. As mentioned above, as of 2005 two-thirds of the University Archives lacked an online descriptive record, and much location and holdings data was outdated. In late 2005, we formulated a new approach to processing and description. The stated goals were to create an online descriptive record for every collection held at the library and to regain basic intellectual control by updating holdings and location information.

A simple Microsoft Access database was in existence, though it had been nearly a decade since the last formal effort to update it. A data entry form was created and populated with data that already existed; records included title, dates of material, extent, and shelving location. Staff then conducted a shelf read of the entire University Archives, updating this very brief descriptive and holdings information. Once completed, reports and search forms were created that assisted in paging and locating material. Reports also assisted in the establishment of processing priorities.

After the initial survey, one staff member was tasked with creating MARC records for each University Archives collection. The staff

member used the database as the foundation for these records but also examined the physical material. These records met DACS's single-level minimum requirements and were created using the library's Integrated Library Management System with a template established at the beginning of the project.

Once creation of the collection-level catalog records was complete (more than 250 records were created in approximately three months), the MARC records were converted to EAD with MarcEdit. The resulting EAD records were loaded into the library's EAD database and increased access to the University Archives collections dramatically. After the creation of the collection-level EAD records, focus shifted to the creation of inventories, which could then simply be attached to the collection-level records. As of 2012, every University Archives collection larger than 2 linear feet was described in a box- or folder-level inventory available through the Princeton finding aids website.

Tools Used: simple Microsoft Access database; Voyager Integrated Library System; MarcEdit

Delivery and Patron Access

Descriptive Data. Data describing material in the Princeton University Archives is delivered to users through both the library EAD finding aids website and through Voyager, the library's integrated library system.

The EAD delivery infrastructure is built on Exist, a native XML database that stores each EAD file and allows for transformation, via XSLT stylesheets, to HTML for viewing on the Web. Exist is compliant with the Xquery standard, which allows for fast and efficient querying of the XML data and a built-in indexing system. As with other XML publishing tools, the implementation of Exist requires either a high degree of technical expertise or significant support from a technology unit.

In the fall of 2011, the Princeton University Library began implementing a discovery layer system called Primo. Primo is a product of Ex Libris Ltd. and is designed to index and deliver descriptive data from a wide variety of sources. The Department of Rare Books and Special Collections EAD working group has been investigating ways to deliver EAD data to users through this system. As of Spring 2012, this work was still in an experimental phase, but we believe we have found

an effective way to utilize this discovery layer as a part of our access system. Our approach is based on the concept of "component records" derived from each EAD component. Complex XSLT transformations are used to create component records that meet DACS's single-level minimum requirements and are enhanced with several additional descriptive elements. These records are validated against a locally developed DTD and finally mapped and transformed to an Ex Libris-specified XML format (PNX) that can be indexed and delivered by Primo.

We plan to fully index and deliver our EAD in Primo by the end of 2012. At that time, we hope to cease using the library's integrated library system as a delivery platform, which will eliminate the need to create MARC records and remove a significant step in our workflow.

The University Archives also maintains a number of legacy databases, describing special format material, such as audiovisual items, artifacts and memorabilia, and files documenting students, faculty, and staff. The data describing this material is stored in separate relational databases. While these databases do allow patrons to search the data, they are, in effect, data silos, and we plan to deliver the data in these databases using our other access systems (our finding aids website and Primo) in the next few years. We are also exploring creating EAC-CPF records for each of the people listed in the student, faculty, and staff databases. These records would also be delivered via Exist.

Tools Used: Exist; Voyager ILS; Primo

Digital Objects. Digital objects are delivered using a variety of methods. Several digitization projects have been conducted in conjunction with the Princeton University Library's Digital Initiatives staff. These projects typically involve the creation of item-level descriptive records in the form of MODS records that meet DACS's single-level minimum requirements, creation of structural metadata in the form of METS records, and delivery of images through the Princeton University Digital Library interface, which is also built on Exist. These projects have typically been labor-intensive and resulted in descriptive records that were not integrated with the rest of our descriptive data.

Since 2008, the Mudd Library has been experimenting with linking digital objects to finding aids. This model relies primarily on PDF files produced on library photocopiers. The resulting PDF files

bundle together the content, usually typescript or handwritten pages, described by each EAD component.

In this model the material described by each EAD component is treated as a single digital object. The principle is that, when creating the description, the archivist already decided that the EAD component represented the fundamental atomic physical unit. We want to rely on decision making by the processing archivist. We create no structural metadata and no item-level descriptive metadata. Practically speaking, common PDF viewers do many of the same things as systems built to take advantage of structural metadata. Each PDF is named using the bar code that was assigned to the physical folder. The only metadata used is the data that was previously available in that EAD component, with the addition of a <unitid>, which is intended to tie the original physical material to the digital object. We use library circulation bar codes for this purpose. A <dao> element is also inserted with the link to the object.

PDF files created for these purposes can be delivered to users easily; at the Mudd Library they are currently managed in the University-supported document management system, Webspace, which is described above.

Tools Used: Exist (for finding aid and metadata delivery); Adobe Acrobat/Reader; Webspace/Zythos or simple Web directory

Crowdsourcing Possibilities. The primary method for gathering user comments and data from finding aids is through the use of a "contact" link present at all times from all finding aids. The library frequently receives corrections and clarifications from users through this method, though it is a very labor-intensive process, requiring users to send e-mail and staff to read the e-mail and act on it. The user commenting feature described in the main body of this module is intended as an improvement to this process.

The University Archives has also implemented a blog entitled "The Reel Mudd," which features digitized content from the Archives and invites users to comment on the posts and provide additional information.

Evaluating Access Systems and User Services

Use Tracking

Use of material is currently tracked with a variety of methods. In-house circulation statistics are kept in a locally designed Microsoft Access database, which allows for the entry of collection call numbers and box numbers. Public Services staff log the call numbers of collections used to answer remote inquiries and that data is also entered into a locally designed database. Similar information is also tracked for photoduplication and image scanning requests. This information is used in a variety of ways, including the establishment of processing and digitization priorities.

As of July 1, 2012, the library has begun testing Aeon, an automated circulation and user registration system. Aeon sends structured data from EAD finding aids to a relational database. It allows sophisticated user tracking and collection of data. The library is also considering implementing Aeon's photoduplication module, which will allow users to request digitization of material directly from finding aids.

Web Analytics

Data on finding aid usage is currently collected using a service called AW stats. At time of writing, the library planned to implement Google Analytics to collect data regarding online finding aid usage. This data will be used to supplement traditional circulation and use statistics and will aid in the implementation of a patron-driven digitization program. For example, the top ten most viewed EAD components will be digitized each month.

Tools Used: local circulation database; Aeon; Google Analytics