

The Archivists' Toolkit

NYU Endeavors to Bring Free Automation to Archives

By Nancy Cricco & Brian Stevens

nancy.cricco@nyu.edu; brian.stevens@nyu.edu

One of the primary responsibilities of an archivist is to hold on to “stuff” (papers, photos, tapes, statues, etc.) and be able to find it when asked. Most archivists deal with quite a bit of “stuff” that is arranged into collections, and the ability to find items in those collections relies on two things: organization and description.

For the uninitiated, archival collections are assemblages of materials of historic or evidential value that have some unifying characteristic, and can range in size from a single folder to hundreds of linear feet. These materials most often comprise individuals' personal papers or an organization's records. A collection's description enables the researcher to determine if what he or she is looking for is in that collection. Organization orders the descriptions in a way that makes the material retrievable. The more material, the more time an archivist must devote to organizing and describing; finding better, more efficient means of doing so has been an ongoing challenge.

Until the 1980s, archives, along with libraries and museums, kept track of their materials on paper. Records of how something was received at an archive (known in the field as an *accession*) and descriptive inventories (known as *finding aids*) were typed up and stored in filing cabinets. Complex, handmade indexing systems were implemented

to help an archivist find material relating to a given subject or person.

And then, technology intervened. Libraries, and then museums, benefited from computers first, primarily because they already had established guidelines for description. In addition, they were usually dealing with a more uniform set of materials than archives: books, paintings, objects, etc. The cataloguing guidelines that

The implication of [the Archivists' Toolkit] is that any archive, regardless of size or resources, will have a significantly greater ability to share its collections with the research community.

had been developed over hundreds of years were transferred to online catalogs relatively easily. A basic database framework was inherent in their manual systems. Unfortunately for archives, there were few established guidelines for description and, where they did exist, there was little adherence to them.

Though archives have existed throughout known history, the profession of archivist is fairly recent

and professional training has only been available since the end of World War II. NYU has one of only a few archival certification programs in the country. In general, archival collections were not organized or described uniformly enough to enable a one-stop software fix for all their problems. Archivists were fairly quick to recognize the problem and, by the early 1990s, organized groups such as the SAA (Society of American Archivists) that began to develop and apply standardization in description and organization. The main vehicle used by archivists for description and organization is the finding aid, so standardization of its format was imperative.

One of these standards was EAD (Encoded Archival Description). Despite the catchy acronym, its practicality and necessity took some time to become widely recognized. NYU, however, was an early implementer of EAD and currently has over 300 collection inventories accessible online with EAD finding aids. EAD enables an archivist to index elements that previously existed statically in typed or word-processed finding aids. At the same time, EAD requires adherence to a standard form with a Document Type Definition (DTD, soon to become a *schema*).

In most cases, researchers find an archival collection by way of an online catalog record that gives the

```

<corpname>New York University</corpname>
<address>
<addressline>
</addressline>
</address>
</repository>
<origination label="Creator" encodinganalog="100">
<persname>Butler, Charles, 1802-1897. </persname>
</origination>
<physloc audience="internal" encodinganalog="852 $z">New York University Archives</physloc>
<abstract label="Abstract" encodinganalog="520 $a">Charles Butler was born in New York in
1802. He was a lawyer whose endeavors included anti-Masonry, New York politics, western
state bond issues and speculation, canal, road and railroad construction, and philanthropy.
The papers of Charles Butler span the years 1817 to 1908, and include correspondence,
business papers, legal documents, ledger books, pamphlets, and maps.</abstract>
</did>
<custodhist id="a16" encodinganalog="561 $a">
<head>Provenance</head>
<p>These papers were donated to the New York University Archives by Charles Butler's
daughter, Emily Ogden Butler, in 1925.</p>
</custodhist>
<accessrestrict id="a14" encodinganalog="506 $a">
<head>Access Restrictions</head>
<p>Open for research without restrictions.</p>
</accessrestrict>
<userrestrict id="a15" encodinganalog="540 $a">
<head>Use Restrictions</head>
<p>Permission to publish materials must be obtained in writing from the:<lb/>
<address>
<addressline>New York University Archives<lb/>
</addressline>
<addressline>Elmer Holmes Bobst Library<lb/>
</addressline>

```

Figure 1. A sample of the XML of Encoded Archival Description (EAD).

researcher a general idea of what a collection contains. This catalog record can then link to a finding aid. The finding aid is intended to be a more comprehensive description of a collection than what can be found or listed in the online catalog record. The finding aid will typically present the researcher with an inventory of the collection's contents that can point the researcher to a specific item or folder that contains the item for which they are searching. When EAD is implemented, finding aids may also be loaded into databases to be searched directly by researchers. Direct searching provides the researcher with additional and more robust search opportunities, well beyond those available through the catalog record.

If two separate repositories have finding aids in EAD, they may also be collected into a larger "union" database because they all speak the same language. This is immensely important for smaller repositories that may not have a significant IT infra-

structure or profile in the research community but hold important collections. If they can include catalog records that point to finding aids in a consortial catalog, their collections are also more likely to be found by the research community.

One hurdle in achieving buy-in for EAD is the fact that it looks like XML (because it is) until it is run through an XSL style sheet (see figures 1 and 2).¹ XML and XSL were, and to some degree continue to be, outside the average archivist's skill set. It therefore requires a significant investment to learn or have someone on your staff learn EAD. In addition, EAD does not hold all the answers. Among its limitations, EAD will only handle description of materials that have been organized; it doesn't address other types of administrative information such as accession location and condition data. An ancillary tool is also needed to search across finding aids in either one's own institution or in a larger consortial setting.

That is where the Archivists' Toolkit enters the picture. In June 2004, New York University Libraries and the University of California, San Diego (UCSD) Libraries, working together with the Five Colleges Libraries, were awarded a grant from The Andrew W. Mellon Foundation to develop a suite of open source software tools for processing and managing archival information beyond what is offered by EAD, while providing EAD as an output.

At present, there is no successful computer-based collection management system tailored to the needs of archival repositories. The Toolkit will address this need by enabling archivists to enter their data into a simple interface and output an EAD finding aid without having to concern themselves with any XML encoding. It will also allow them to manage other parts of their operation, such as keeping track of the location of materials and who donated them, how many researchers are using what materials, what requires pres-

1. Extensible Markup Language (XML) is a programming language for the Web that allows for customized tags, enabling richer definition and easier transmission of data. Extensible Style Language, or XSL, is a specification that, similar to the concept of templates, allows programmers to dynamically apply a single style to multiple XML (or HTML) documents.


<p>TABLE OF CONTENTS</p> <p>Descriptive Summary</p> <p>Historical/Biographical Note</p> <p>Scope and Content Note</p> <p>Arrangement</p> <p>Restrictions</p> <p>Access Points</p> <p>Related Material</p> <p>Separated Material</p> <p>Administrative Information</p> <p>Container List</p>	 <p>Guide to the Charles Butler Papers 1817-1908 (Bulk 1837-1882)</p> <p>New York University Archives Elmer Holmes Bobst Library 70 Washington Square South New York, NY 10012</p> <p>Phone: (212) 998-2646 Fax: (212) 995-4070 E-mail: nancy.cricco@nyu.edu</p> <p>© 2003 New York University Archives. All rights reserved.</p> <p>New York University Libraries, Publisher</p> <p>Processed by Cynthia Siu and Geri Solomon</p> <p><small>Machine-readable finding aid derived from a EAD v.1 finding aid dated: 2000. Machine-readable finding aid created by Katie Senft . Description is in English.</small></p>
---	---

Figure 2. The EAD XML once it has passed through XSL, creating an HTML document.

ervation attention, and even whether an acknowledgement has been sent to the donor of a collection. All these functions are normally handled by separate software tools that have little or no integration, which results in a highly segmented and redundant workflow that is inefficient and costly.

To ensure the development of a truly comprehensive software tool, the Archivists' Toolkit will be created with the input of seventeen archival repositories that represent a broad range of workflows, sizes, materials, staffing, and resources. The repositories participating in the project are:

- **In New York City** – The American Museum of Natural History, The Brooklyn Museum of Art, Carnegie Hall Archive, The Center for Jewish History, Manhattan College, NYU's Fales Library & Special Collections, University Archives, and Tamiment Library & Wagner Labor Archive
- **In western Massachusetts** – Amherst College Archives and Special Collections, Hampshire

College Archives, Mount Holyoke College Archives and Special Collections, Smith College Archives, Sophia Smith Collection, and the University of Massachusetts Amherst Special Collections and Archives

- **In southern California** – UCSD's Mandeville Special Collections Library and Scripps Institution of Oceanography Archives

The project has received funding through June 2006, and its management is based at the UCSD Libraries. The programming team is located at the NYU Libraries, with specifications being developed at UCSD, the Five Colleges, and NYU. The specific features of the application will be vetted and tested with the staff at the above fifteen archival repositories in the San Diego, New York City, and Amherst areas.

Because the Archivists' Toolkit will be free and open source, anyone will be able to obtain the software and programmers will be able to tailor the software to the needs of individual repositories. If the Toolkit

is successful, it will decrease the time and costs associated with archival processing and promote the standardization of archival information. This will benefit researchers by increasing the amount of material processed, making it possible for more of a repository's budget to be shifted from archival description to collection development or public service. In addition, it will facilitate the development of more sophisticated and granular union catalogs of archival information. The implication of this is that any archive, regardless of size or resources, will have a significantly greater ability to share its collections with the research community.

For more information about the Archivists' Toolkit, please visit <http://clio.bobst.nyu.edu/toolkit/>.

Nancy Cricco is University Archivist at the New York University Archives; Brian Stevens specializes in Encoded Archival Description at the NYU Archives and is the Project Archivist and Designer for the Archivists' Toolkit.