The Archivists’ Toolkit: An Integrated System for Describing and Managing Archival Resources

Kelcy Shepherd, Bradley D. Westbrook, Lee Mandell, Brian Stevens, Jason Varghese

Introduction

The Archivists’ Toolkit (AT) project is an effort to create an open source software system that will provide archivists with a single integrated tool for recording, storing, analyzing, and distributing information related to the description and management of archival materials. The project involves the University of California San Diego (UCSD), New York University (NYU), and the Five Colleges, Inc., and has been funded by The Andrew W. Mellon Foundation. This article discusses the issues that motivated the project, features of the Archivists’ Toolkit software application, archival functions to be supported by the Archivists’ Toolkit, as well as participants in the project.

The Need for an Integrated Archival Management System

Recording information about archival resources is an ongoing, iterative process. The archivist may begin registering basic data related to the papers or records as soon as she is in contact with the source (e.g., donor, dealer, office of origin). Upon acquiring the materials the archivist records additional information during the accessioning process. The purpose of accessioning is to gain basic physical and intellectual control over the materials by recording information about the accession transaction: when the materials were received, from whom, and how much was acquired. The accession record may also include preliminary information about the source(s) and the creator(s) of the materials, as well as the scope, content, physical condition of the collection and its location in the repository.

Once the archival collection is arranged and transferred into acid-neutral files and boxes, the archivist will create a more detailed description. The full description of the materials may be multi-level, encompassing information about the collection as a whole and information about record series, files, and/or items within the collection. Throughout these activities, the archivist must also manage the physical location of the collection and track workflow for the acquisition and processing of the collection.

Traditionally, archivists in the United States have used multiple, disparate systems to support the iterative process of managing and describing the collections in their care. Early analog systems for recording and using donor, accessions, description, and location information were replaced with discrete electronic systems, or a combination of analog and digital systems. It is not uncommon for archivists to use an internal database to record accessions data, word processed documents for creation and distribution of archival descriptions, and a spreadsheet or card files for tracking locations. Obviously, it is not easy to carry and reuse data captured at one point in the archival workflow to a later phase, using systems that are not
integrated.

The difficulties of using multiple systems to manage archival information have intensified as new standards and technologies have emerged to enable distribution and exchange of archival descriptions. The advent of Machine-Readable Cataloging for Archival and Manuscript Control (MARC-AMC) in 1982 allowed archivists, especially those in library settings, to create abbreviated collection-level descriptions and make them available to researchers through their local library catalogs and through national union catalogs such as Research Libraries Information Network (RLIN) and the Online Computer Library Center (OCLC) union catalog. It also added another task to the archival workflow.

In the early 1990s U.S. archivists began using the Internet to distribute finding aids for their collections. Unlike a discrete bibliographic catalog record, which can only accommodate a brief single-level description, archival finding aids are often multi-level descriptions. Not only do these multi-level finding aids contain a more thorough description of the context and contents of the archival resource, they also provide descriptive information about each of the hierarchical groupings within that resource (e.g. descriptions of records series, listings of folders comprising the collection, etc.). To be sure, the ability to publish entire finding aids online allowed archivists to provide researchers with richer, more detailed descriptions than those included in library catalog records.

Archivists initially used WAIS, Gopher and then HTML to mark up and deliver finding aids, but their limitations for facilitating searching and navigating online finding aids quickly became apparent. Furthermore, HTML encoding, while it improved the presentation of finding aids, did not help to promote consistent application of encoded data elements within or across repositories. The desire for a community-based standard encoding structure for archival finding aids led to the development of a Standard Generalized Markup Language Document Type Definition (SGML DTD) specifically for encoding archival collection descriptions and facilitating their publication online. This DTD, known as Encoded Archival Description (EAD), allows archivists to represent the hierarchical structure inherent in archival collections in an encoded format and then use this encoding for searching and navigating through a finding aid or a group of finding aids. EAD, which has evolved into an Extensible Markup Language (XML) standard, has been adopted widely in the United States, as well as in Canada, Australia, the United Kingdom, France, Germany, and the Netherlands. But, at least in the United States, EAD is implemented almost without exception as an added activity, and not as a byproduct of routine descriptive work.

Not long after U.S. archivists began publishing finding aids online, they were also began exploring digitization of archival materials - scanning images, transcribing and encoding texts, and capturing sound and moving images in digital form. The ability to publish digital surrogates of unique materials online, globally accessible to anyone with an Internet connection, opened new avenues for researchers and introduced new users to archival resources. EAD provides elements for linking to digital surrogates of archival materials, not to mention born digital archival materials, within an encoded finding aid. In addition, archivists are also using other metadata schemas developed within the digital library community to enhance access to their digital surrogates.

MARC, EAD, and other metadata standards, combined with the technology to create and distribute digital surrogates of selected archival materials, have added multiple complex layers to the work of archivists. Such standards also present opportunities to increase access to information about archival materials, to publish surrogates of resources once available only within the archives itself, and to promote traditional and new uses of archives. Archivists are eager to capitalize on these opportunities. But, most would welcome better tools for doing so - tools that allow them to record information at the point of donor contact and carry it forward when describing materials, to focus on training staff to use a single system with a user-friendly interface rather than teaching them the mechanics of MARC coding and XML tagging for diverse schemas, and to use the same data to produce a variety of administrative reports and descriptive outputs. The Archivists' Toolkit pro-
# The Archivists’ Toolkit: An Integrated System for Describing and Managing Archival Resources

## Archival Functions Supported in The Archivists’ Toolkit, Version 1.0

<table>
<thead>
<tr>
<th>Archival Activity</th>
<th>Purpose</th>
<th>New Data Recorded*</th>
<th>Existing Data Utilized or Refined</th>
<th>Reports and Other Tools Produced*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Contact with source of materials to be acquired (such as the donor, office of origin, dealer, etc.)</td>
<td>Identify and obtain materials for addition to the repository’s holdings</td>
<td>Source contact information, Notes tracking conversations, describing materials to be acquired, etc.</td>
<td>Name of source</td>
<td>Source contact list</td>
</tr>
<tr>
<td><strong>Accessioning</strong></td>
<td>Provide basic physical and intellectual control over incoming material. Data recorded is generally preliminary in nature, and will be enhanced when the materials are described.</td>
<td>Accession Number, Accession Date, Title, Date of materials, Extent, Description, Condition, Acknowledgement Date, Agreement Sent, Agreement Received, Intellectual property rights, Access restrictions</td>
<td>Name of source, Name of creator, Subjects</td>
<td>Accession record, Acquisition statistics, Workflow tracking lists for activities such as obtaining deeds of gift, acknowledging receipt of materials, etc.</td>
</tr>
<tr>
<td><strong>Arrangement and description (commonly referred to as &quot;archival processing&quot;)</strong></td>
<td>Preserve, organize, and describe materials in order to make them accessible for use by patrons</td>
<td>Resource Identifier, Level Label, Resource Dates, Extent, Notes (Scope and contents, accruals, appraisal, arrangement, conditions governing use, conditions governing access, custodial history, physical description, immediate source of acquisition, language of materials, processing information, related archival materials, etc.)</td>
<td>Title, Name of source, Name of creator, Subjects</td>
<td>Finding aids, Catalog records, Metadata records</td>
</tr>
<tr>
<td><strong>Name authorities management (may occur concurrently with accessioning or description, or as a separate workflow)</strong></td>
<td>Create and manage vocabularies of authorized personal, corporate, and family names that represent sources, creators, and subjects of archival materials</td>
<td>Name type, Name source, Name rules, Name (broken into appropriate sub-elements for each type, for example last name, rest of name, fuller form of name, etc.), Biographical or administrative history, Non-preferred form of name</td>
<td>Name of source</td>
<td>Authority records (such as MADS, EAC)</td>
</tr>
<tr>
<td><strong>Subject authorities management (may occur concurrently with accessioning or description, or as a separate workflow)</strong></td>
<td>Create and manage vocabularies of uniform title, topical, geographic, genre/form, occupation, and function terms</td>
<td>Subject term type, Subject term, Subject term source, Subject scope note</td>
<td>Subject guide for repository holdings</td>
<td></td>
</tr>
<tr>
<td><strong>Location management (occurs when materials are accessioned and again if they are moved)</strong></td>
<td>Maintain physical control over holdings, allow for retrieval of materials</td>
<td>Location of a container (box, carton, flat file, etc.)</td>
<td>Container information from description or accession record</td>
<td>Shelf listings, List of locations for a collection</td>
</tr>
</tbody>
</table>

* Lists of New Data Recorded and Reports and Other Tools Produced are not intended to be exhaustive, but merely to illustrate some basic examples for each activity.
ject is an effort to provide one such tool.

Features of the Archivists' Toolkit

The Archivists' Toolkit is to be an open source database application that will support management and description of archival materials, increase archival processing efficiency, lower the costs of archival description in several ways, and promote standardization for description of archival resources. The application is designed to support essential functions of the record cycle, to promote community descriptive standards, to allow for increased efficiency in archival processing, and, very importantly, to be customizable to local work settings.

The Archivists' Toolkit will:

- Satisfy and integrate key functions in the archival descriptive cycle, including accessioning, registering source information for archival resources, creating and applying name and topical headings to resources, managing resource locations and producing finding aids, collection guides, complex digital objects, and various administrative reports.

- Allow archival repositories to ingest existing metadata about collections in the form of EAD finding aids, MARC records, and electronic accession information into a centralized, integrated database, and store and process it along with newly created information about digital assets, accessions, and sources.

- Be deployable in a range of archival repositories from historical societies, college archives, museum archives, commercial archives, and other archives specializing in non-textual materials (sound or video archives, for example) on the one hand, and, on the other, in archives with a single staff member or in multi-repository consortia such as the Five Colleges, Inc., or the Online Archive of California.

The Archivists' Toolkit is being designed for flexibility and ease of use. Elements of the application interface will be customizable, with individual repositories being able to choose which data elements to display in an input template, how to label those elements, and the order in which the elements should appear. Thorough, user-friendly documentation, including installation guidelines, complete user manual, data dictionary, glossary, and in-context help and error messages will be provided.

Archival Functions Supported in the Archivists' Toolkit

The core functions that will be supported in the version 1.0 of the Archivists' Toolkit are accessioning and deaccessioning, description of archival resources and digital objects (including application of name and subject headings), and location tracking. The AT application includes an additional functional area, for managing the application and defining projects, which is not covered in detail here. A fuller description of each functional area, complete with tasks lists, data requirements, flow diagrams, and element definitions is available in the software specification.

Accessioning

The application will provide for the creation and editing of accession records. At its minimum, the accession record represents the accession transaction, but additional data elements are provided for description of the accession and for basic collection management information. Among these data elements are a summary of the content, note on physical condition, a preliminary description of the resource and its condition, whether intellectual rights have been transferred or if restrictions constrain use of the resource, and whether acquisition agreements have been acquired and acknowledgement letters sent.

Deaccessioning

The application will also support tracking the removal of materials from accessions. Repositories will be able to record what materials were removed, how much was removed, why it was removed, what was done with the removed materials, and if the source of the materials was notified.

Description

The description functional area constitutes the largest part of the application. This functional area complies with General International Standard Archival Description (ISAD (G)) and Describing
The Archivists’ Toolkit: An Integrated System for Describing and Managing Archival Resources

Archives: A Content Standard (DACS) and will support the single and multi-level description of archival resources and of digital objects drawn from those resources. Repository staff will be able to import and, if necessary, then edit information from accession records. The application will also provide 27 descriptive notes, each of which can be applied multiple times to every component level. In addition, it will be possible to describe multiple manifestations of an information object at the same component level. Thus, an original document, a microfilm version, and a digital image version can all be described at the same component level.

Names

Name records in the AT are based on the International Standard Archival Authority Record for Corporate Bodies, Persons, and Families (ISAAR (CPF)) and are devised to facilitate the output of name authority records compliant with the emerging Encoded Archival Context (EAC) standard. Repositories will be able to create name records for persons, families, or organizations and link them to accession and description records to indicate the agent who created the materials, or who was the source of the materials, or who is the subject of the materials.

Subjects

Descriptors for content or format characteristics are recorded in subject records in the AT, which can then be applied to accession or description records and output, for instance, as part of EAD and MARC records. The subject term element is designed primarily to accommodate simple terms or phrases such as “Boats” or “Boats and boating” and qualified terms such as “Boats and boating – Law and legislation.” The record also permits for recording the type of subject term (topical, geographic, genre, etc.), the authority source, and a scope statement for the term that can be consulted during data entry.

Location

An important function for any integrated archival information system is to record and represent both permanent and temporary locations of archival resources. This is vital for tracking such materials, which often migrate from permanent shelf locations to reading room locations, to exhibition locations, to conservation laboratories and so forth. In its first release, the AT will only support recording the permanent location of archival materials, that is, the location at which archival materials are normally located when they are not in use for some purpose or undergoing some treatment. It will be possible to create location records at the level of an archival collection or at the level of a component part such as a box, a microfilm reel, or a URL for a digital file.

Project Participants

The Archivists’ Toolkit Project Team includes a project manager based at UCSD; a design team manager, a programmer, and an archives analyst based at NYU; and an archives analyst at the Five Colleges and two Principal Investigators, one at UCSD and the other at NYU. In addition, seventeen archival repositories have agreed to assist the Project Team with developing the functional requirements for the AT and, when the prototype becomes available, with testing the usability of the application’s interface. The seventeen partners are:

Five Colleges, Inc.
Amherst College Archives and Special Collections
Hampshire College Archives
Mount Holyoke College Archives and Special Collections
Smith College Archives
Sophia Smith Collection, Smith College
Special Collections and University Archives, University of Massachusetts Amherst

Participating New England Area Archives
The Edmund S. Muskie Archives and Special Collections Library, Bates College

New York University
Fales Library, NYU
NYU University Archives
Tamiment Library, NYU

Participating New York City Archives
The American Museum of Natural History
The Brooklyn Museum of Art
Finally, an Advisory Board composed of six expert archivists, archival educators, and digital library administrators serves the Project Team by providing guidance on the scope of the AT’s initial development phase, appropriateness of the proposed functional requirements, overall application design, and marketing. The AT Advisory Board members are:

Robin Chandler, California Digital Library
Michael Fox, Minnesota Historical Society
Merrilee Proffitt, Research Libraries Group
Richard V. Szary, Yale University
Guenter Waibel, Research Libraries Group
Beth YakeL University of Michigan

Presentations at regional and national conferences also provided a venue for the Project Team to elicit feedback from additional members of the archives, library, and museum communities.

Conclusion

Project staff completed a detailed functional specification for the Archivists’ Toolkit in the fall of 2005, and an abridged version made available online for public review (http://archiviststoolkit.org/ATspecification/index.html). A prototype of the AT, based on the specification, is currently in development and functional testing phases. User acceptance testing for the system will take place during the summer of 2006, and the Project Team plans to release version 1.0 of the application in the fall. Further information about the project is available on the Archivists’ Toolkit web site (http://www.archiviststoolkit.org/)

The functional areas that are being built for the first release do not, of course, represent all the functions of the archival life cycle. The Project Team expects functionality of the application will be substantially expanded in subsequent development cycles. Even in its initial release, the Archivists’ Toolkit will be a single system supporting multiple archival functions, functions that archivists now fulfill through the time-consuming process of learning, using, and managing many different tools.


2) Prior to MARC format integration, MARC-AMC was the MARC format used specifically for describing archival and manuscript collections. For information about MARC 21, see http://www.loc.gov/marc/. Accessed March 3, 2006.


6) An abridged version of the application specification, which provides a complete description of the application, is available on the AT project site: http://archiviststoolkit.org/ATspecification/index.html. Accessed March 24, 2006.


8) Describing Archives: A Content Standard (DACS) is the United States standard for the description of archival materials, approved by the Society of
