WHAT IS ENCODED ARCHIVAL DESCRIPTION (EAD)?

EAD is an XML/SGML-based DTD for electronic finding aids.

This concise, but non-illuminating definition of EAD can be greatly improved upon with a brief dissection of its primary elements:

A finding aid is any tool that is used for resource discovery and use of archival materials. A finding aid is the primary method used by archivists for conducting archival description – the activity of creating, exchanging and publishing information about sets of historical materials in a collection. A finding aid can assume many forms, the most common being inventories and registers. They can describe collections at multiple levels of detail ranging from individual items to records groups and series. A key feature of traditional, paper-based finding aids is their columnar format, which supports multiple access views of a collection. One column may provide physical storage location information, while another provides descriptive data:

A traditional finding aid, displaying columnar formatting. A bit blurry, sorry!
EAD (Encoded Archival Description) is a metadata standard developed in the mid-1990’s for electronically encoding the archival description information contained in finding aids. An established set of rules for electronically encoding finding aid documents, it is specifically a Document Type Definition (DTD); that is, a formal declaration defining the set of EAD markup tags used to identify elements of a document. Additionally, it provides grammatical rules for relationships between elements, sequence of elements, attributes and values of elements. Standardized and platform-independent, a DTD facilitates resource discovery, information and display, and data exchange between different organizations in a distributed network environment.

The semantic tags and rules contained in the DTD are expressed and executed via the standard syntax provided by either Extensible Markup Language (XML) or its broader counterpart, Standard Generalized Markup Language (SGML). These are both programming specifications used for formatting structured electronic documents. XML is a subset of SGML, specifically designed for the web environment.

Thus, EAD is a particular DTD, used in the electronic formatting of archival finding aids. It is a set of formal definitions for identifying and describing elements of a finding aid. It provides the discipline-specific, semantic ‘ground rules’, whereas XML (or less often, SGML) provides a universal syntactic structure needed to express and communicate electronic information between organizations.

**EAD PURPOSE AND CHARACTERISTICS**

By providing structural uniformity of finding aid information, EAD can be used by archives, museums and libraries to create robust electronic descriptions of their collections. These descriptions support customized and flexible display, navigation, search, retrieval and exchange of finding aid information, across organizations. Publication of EAD documents facilitates research activity by promoting access and resource discovery of an institution’s holdings.

EAD was developed by archivists, for archivists. As a result of this in-depth understanding of the purpose and use of traditional finding aids. EAD documents bear many of the same important characteristics of textual finding aids:

- EAD documents are structured **hierarchically**, with nested levels of increasingly detailed data.
- Each successive hierarchical level inherits information, parent to child.
- EAD documents are **recursive**, with many of the same data elements repeated at each hierarchical level.

These key features enable a single finding aid to fully describe a collection from various perspectives – from broad, overarching terms to descriptions of individual items.
Unfortunately, the creation of EAD required some compromise. Whereas the columnar format of traditional finding aids reflects both intellectual and physical facets of a collection, EAD cannot effectively support this dual structure. The design committee elected to focus on the intellectual arrangement and description of archival materials; physical location information is still included, but is not the prominent perspective for describing archival material.

EAD ORIGINS

EAD grew out of a recognized need for electronic finding aids in an increasingly networked environment. Machine-readable finding aid formats like the AMC-MARC standard (Archival and Manuscript Control) had been available as early as 1983, but were flat-structure formats that failed to provide the hierarchical views essential to complex finding aids. Beginning in 1993, initial project development was led by Daniel Pitti of the University of California – Berkeley.

In 1995, in a one-week meeting at Bentley Historical Library in Ann Arbor, MI, Pitti’s prototype was reviewed and revised by a design team of seven archivists and one SGML expert, and the initial structure of EAD was created.

An alpha version of EAD was released to the public in 1996, and was rapidly accepted by the archival community. The first EAD documents were created within just months of the initial release of the standard. A beta version was released in 1997, followed by first official version 1.0 release in 1998. This latter version incorporated compatibility with the emerging XML standard for data exchange on the web.

The most recent version of EAD was released in 2002. Maintenance and continuing development of the standard is now jointly conducted by the Network Development & MARC Standards Office of the Library of Congress, and the Society of American Archivists. Full EAD. Full specifications of the 2002 release may be viewed at http://www.loc.gov/ead/, the official EAD web site.

HIGH-LEVEL EAD ELEMENTS

A full discussion of EAD elements is well beyond the scope of this paper; this section introduces only the primary, high-level features. A EAD document is initially broken into two segments, with three major tags (highlighted):

1. Information about the finding aid:

   `<eadheader>`

   This element provides basic identification data. Sub-elements must be listed in a prescribed sequence:
Carol Smith, p. 4

<eadid> Unique identification number
<filedesc> Bibliographic data broken into numerous sub-elements for title, author, edition, series, publisher, etc..
<profiledesc> Creation data broken into two further sub-elements for language and creator.
<revisiondesc> Records any revisions to EAD document.

<frontmatter> Provides optional elements that for creating a finding aid title page. Sub-elements may be listed in any preferred sequence:

<titlepage> Broken into many of the same sub-elements as for <filedesc>
<div> A general tag for other title page items such as prefaces, dedications, etc...

2. Information about the archival materials:

<archdesc> Provides a hierarchical description of a collection of archival materials, in succeedingly finer detail. Sub-element tags are extensive. A small, representative sample:

<did> "Descriptive ID". Basic description info broken into numerous sub-elements for container, physical location, repository, etc..
<add> Additional descriptive data. Sub-elements include tags for related materials, index, notes, etc..
<admininfo> Administrative data. Sub-elements include tags for access restrictions, custodial history, appraisals, etc..

A DTD document can be endlessly nested for complete, detailed description of resources contained in an archival collection. Excellent sample code can be viewed at http://www.loc.gov/ead/ag/agappe.html.

SIGNIFICANCE OF EAD

EAD has enjoyed great success domestically as well as internationally. Rapidly accepted by the U.S. archival community, EAD is now the standard for machine-readable finding aids. EAD was specifically designed to comply with ISAD(G), the pre-existing general international standard for archival description. As a result, it gained broad, early international adoption, including acceptance by non-English speaking nations (Sweet, 2001).
Despite its success as a standard, EAD faces some challenges. Because of its rich descriptive capabilities, EAD is perceived as complex, and difficult to implement. EAD provides a great deal of flexibility, which has resulted in a lack of uniform implementation practices across institutions (Tatem, 1998). Conversion of existing electronic finding aids from other standards to EAD such as MARC has proven to be slow and costly, as has production of new EAD documents. Barriers to international adoption, arising from differing archival practices, still need to be addressed (Sweet, 2001). All these issues are further compounded by a relatively slow adoption of technology by the archival community, and by a distinct lack of user studies, impeding understanding of how EAD is understood and used (Reiner, & Sweet, 2005, abstract, p. 296).

Adoption of EAD will continue to grow, requiring resolution of these issues. Expanded development and use of EAD-related technology is seen as a way to address many challenges. Working improvements include:

- Simplified production rules, to promote uniform coding practices (e.g., EAD Cookbook).
- Automated production and conversion tools.
- Improved EAD search/retrieval software solutions.
- Improved crosswalk capability with other metadata standards such as ARM-MARC and ISAD(G).

Proponents of EAD see possible applications of the standard beyond finding aids, including its incorporation into union catalogs, and its use as a creator and record descriptor, for general bibliographic control purposes. Some of these visions are now coming into fruition; for example, the Northwest Digital Archives (NWDA) has contributed approximately 2,300 EAD-based finding aids to a union catalog of western archival materials (Cornish, 2004, p. 181).

**PERSONAL EXPERIENCE?**

I regret that I can claim no working experience with EAD or any other major metadata standards listed in this week’s assignment.

As an avid generational history researcher, however, I am an experienced user ofGEDCOM (GEnealogy Data COMmunication), a widely supported, privately developed specification for the exchange and formatting of genealogical data. Newer specifications like GedML can express the GEDCOM model in XML syntax, and are gaining slow acceptance in the genealogical community.
REFERENCES


EAD IMPLEMENTATION – CASE STUDIES

