



Metadata: Essential Standards for Management of Digital Libraries

ALI Digital Library Workshop
Linda Cantara, Metadata Librarian
Indiana University, Bloomington
2 October 2003



What is Metadata?

- “data about data”
- describes various aspects of a digital file or group of files
- identifies the parts of a digital object and documents their content, location, relations, structure, and functionality



Metadata “Buzzwords”

- Interoperability
 - the ability of software and hardware on different machines from different vendors to share data
- Modularity
 - constructed with standardized units or dimensions for flexibility and variety in use
- Extensibility
 - capable of being increased in scope or range



XML:

eXtensible Markup Language

- subset of Standard Generalized Markup Language (SGML) -- international standard ISO 8879
- metalanguage for designing markup schemas for electronic resources
- facilitates definition, transmission, validation, and interpretation of data between applications and between organizations



Why XML?

- separates document structure from document presentation
- nonproprietary human and machine-readable electronic format
- facilitates indexing content for browsing and searching on-line



What is a DTD?

- A Document Type Definition (DTD) provides the rules for a specific implementation of XML. The DTD states what elements (also called tags) and attributes (additional information about an element) may be used, where and how each tag may be used, and which elements are allowed to occur within other elements (also called nesting).



Types of Metadata

- descriptive metadata
- administrative metadata
 - technical metadata
 - rights metadata
 - source metadata
 - digital provenance metadata
- preservation metadata



Descriptive Metadata

facilitates

- description of the source object used to create the surrogate digital object as well as a collection of objects
- discovery and identification of the collection or object
- access to the collection or object
- browsing and searching



Administrative Metadata [1]

- technical metadata: information regarding a file's creation, format, and use characteristics
- rights metadata: copyright and licensing information
- source metadata: descriptive and administrative metadata regarding the analog source from which a digital library object derives



Administrative Metadata [2]

- digital provenance metadata: information regarding source/destination relationships between files, such as
 - master/derivative images: archival tiffs, hi-resolution jpegs, low-resolution thumbnails
 - information regarding migrations/transformations of a file from the original digitization of an artifact to its current incarnation as a digital library object



Preservation Metadata

All the above is necessary to ensure long-term access, interoperable transmission and exchange of digital resources.



Metadata Standards

- ensure persistent user access to metadata
- facilitate metadata interchange
- support long-term metadata viability



Types of Metadata Standards [1]

- data structure standards
 - establish data elements and attributes
 - "buckets" of information



Data Structure Standards

- MACHine-Readable Cataloging (MARC)
- Text Encoding Initiative (TEI)
- Encoded Archival Description (EAD)
- Dublin Core (DC)
- Visual Resource Association Core Categories (VRA Core 3.0)
- Metadata Object Description Schema (MODS)
- NISO Metadata for Images in XML Schema (MIX)



Types of Metadata Standards [2]

- data communication standards
 - means of storing, manipulating, and displaying metadata
 - facilitate interchange of information



Data Communication Standards

- XML
- MARC
- MARCXML
- Metadata Encoding and Transmission Standard (METS)



Types of Metadata Standards [3]

- data content standards
 - govern order and syntax of information included in metadata “buckets”



Data Content Standards

- Anglo-American Cataloging Rules (AACR2)
- Archives, Personal Papers, and Manuscripts (APPM)
- Graphic Materials: Rules for Describing Original Items and Historical Collections (GIHC)



Types of Metadata Standards [4]

- data value standards
 - govern choice and form of data within elements
 - controlled vocabularies and ontologies



Data Value Standards

- Library of Congress Subject Headings (LCSH)
- Library of Congress Authority Files (LCAF)
- Art & Architecture Thesaurus (AAT)
- Getty Thesaurus of Geographic Names (TGN)
- Union List of Artist Names (ULAN)
- Thesauri of Graphic Materials (TGM I and TGM II)



Selecting Metadata Standards

project analysis

- nature of project (text, images, audio, video, all-of-the-above)
- anticipated lifespan (temporary, short-term, long-term)
- choice of software and interface for display and searching



Metadata Crosswalks

- map the equivalent relationships between two or more metadata formats
- promote interoperability
- facilitate metadata harvesting (i.e., OAI-PMH)



EAD:

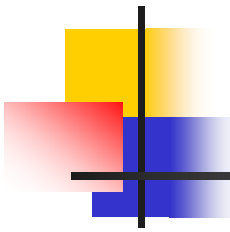
Encoded Archival Description

- standard for electronic encoding of finding aids for archival and manuscript collections
- implementation of Extensible Markup Language (XML)
- maintained in the Network Development and MARC Standards Office of the Library of Congress (LC) in partnership with the Society of American Archivists



History of EAD

- 1993: initiated by the University of California, Berkeley Library under direction of Daniel Pitti
- August 1998: version 1.0 Document Type Definition (SGML DTD)
- December 2002: version 2002 SGML/XML DTD



Benefits of an EAD finding aid (vs. MS Word or HTML)

- documents the interrelated descriptive information of an archival finding aid
- preserves the hierarchical relationships existing between levels of description
- represents descriptive information that is inherited by one hierarchical level from another
- supports element-specific indexing and retrieval of descriptive information



Why create EAD finding aids?

- on-line publication of guides and inventories to archival and manuscript collections
- facilitate searching and browsing content of collections
- may provide access to digital images and/or transcriptions of items in a collection



How to create an EAD finding aid

- design template for local implementation
- encode data from existing finding aids (MS Word documents, HTML files, etc.)
 - copy-and-paste
 - rekey
- create original finding aids from existing collection inventories



Publishing an EAD finding aid

- transform XML EAD to HTML for on-line display
 - using XSL stylesheet
 - using configurable on-line XML publishing software
- create an index of XML tags for browsing and searching
- optional: link descriptions to digital images, text transcriptions, audio files, etc. of the item described



Examples

- [Indiana University Finding Aids](#)
- [Five College Archives & Manuscripts Collection](#)
- [Washington Research Libraries Consortium](#)
- [Kentuckiana Digital Library](#)



More Information...

- EAD Home Page

<http://www.loc.gov/ead>

- EAD References

<http://www.iu.edu/~lmc/ead-refs.htm>