

Part 1: METADATA STANDARDS

Metadata Standards Identified for Use Include:

Descriptive Encoding Standards

- **Core Standards – Archival Description**
 - Dublin Core (DC)
 - Encoded Archival Description (EAD)
- **Supporting Standards – Bibliographic Description**
 - MACHine-Readable Cataloging for the 21st Century (MARC 21)
 - Metadata Object Description Schema (MODS)
- **Supporting Standards – Visual Resources Description**
 - Categories for Description of Works of Art (CDWA) / CDWA Lite
 - Visual Resources Association Core Data Standard (VRA)

Data Content Standards & Rules

- Describing Archives: a Content Standard (DACS)
- Anglo-American Cataloguing Rules (AACR2)
- Cataloging Cultural Objects (CCO)

Data Value Authorities

- Art and Architecture Thesaurus (AAT)
- Union List of Artist Names (ULAN)
- Thesaurus of Graphic Names (TGM)
- Subject Headings (LCSH)
- Thesaurus for Graphic Materials: Subject Terms (TGMI)
- Thesaurus for Graphic Materials: Genre and Physical Characteristic Terms (TGMII)

Data Exchange / Resource Discovery Standards

- Metadata Encoding Transmission Standard (METS)
- Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH)
- XML Schema for the Sitemaps & Sitemap Index Files

APPENDIX F: Standards, Best Practices, and Tools

Descriptive Encoding Standards

Core Standards – Archival Description

- **Dublin Core (DC)** – <http://dublincore.org/>
Set of 15 descriptive semantic definitions that provide an easy to use semantic vocabulary for describing the "core" information properties, such as "Description" and "Creator" and "Date".
 - **Encoded Archival Description (EAD)** – <http://www.loc.gov/ead/>
Encoding standard for finding aids including inventories, registers, indexes, and other documents created by archives, libraries, museums, and manuscript repositories to support the use of special collections.
-

Supporting Standards – Bibliographic Description

- **MAchine-Readable Cataloging for the 21st Century (MARC 21 / MARCXML)** – <http://www.loc.gov/standards/marcxml/>
Simple XML schema used to generate full MARC records are needed or act as a carrier to enable MARC records to go through further transformations such as to Dublin Core and/or processes such as validation.
 - **Metadata Object Description Schema (MODS)** – <http://www.loc.gov/standards/mods/>
Bibliographic element set expressed as a human readable XML schema, developed by LOC as a compromise between Dublin Core's simplicity and the complexity of MARC21.
-

Supporting Standards – Visual Resources Description

- **Categories for Description of Works of Art (CDWA) / CDWA Lite** – http://www.getty.edu/research/conducting_research/standards/cdwa.html
CDWA describes the content of art databases through a framework for describing and accessing information about works of art, architecture, other material culture, groups and collections of works, and related images, and includes 512 categories and subcategories, as well as basic guidelines for cataloging. **CDWA Lite** is a small subset of core CDWA categories that represent the minimum necessary to identify and describe a work.
 - **Visual Resources Association Core Data Standard (VRA)** – <http://www.vraweb.org/vracore3.htm>
The VRA Core provides a template designed for a visual resources collection, such as art and architectural works, and is based on an element set contains two groupings of elements, the Work Description Categories (19 elements), and the Visual Document Description Categories (9 elements). Like CDWA Lite, VRA Core 4.0 assumes the use of CCO for cataloging guidelines.
-

APPENDIX F: Standards, Best Practices, and Tools

Data Content Standards & Rules

- **Describing Archives: a Content Standard (DACS)** – <http://www.archivists.org/catalog/pubDetail.asp?objectID=1279>
DACS is a set of rules for describing archives, personal papers, and manuscript collections that can be applied to all types of materials. It is the US implementation of international archival description standards such as ISAD(G).
 - **Anglo-American Cataloguing Rules (AACR2)** – <http://www.aacr2.org/>
AACR2 contains rules of description for library materials. These rules can be used as content guidelines for describing digital objects, but is designed for records appearing in library catalogues.
 - **Cataloging Cultural Objects (CCO)** – <http://www.vraweb.org/ccoweb/cco/index.html>
Cataloging Cultural Objects: A Guide to Describing Cultural Works and Their Images (CCO) by Murtha Baca (Chicago: American Library Association, 2006) is a manual for describing, documenting, and cataloging cultural works and their visual surrogates and focuses on art and architecture, including paintings, sculpture, prints, manuscripts, photographs, built works, installations, and other visual media. CCO deals with information related to a subset of the CDWA Categories and the VRA Core Categories.
 - **Metadata Object Description Schema (MODS)** – <http://www.loc.gov/standards/mods/v3/mods-userguide.html> An XML schema for descriptive metadata compatible with the MARC 21 bibliographic format that uses a smaller core subset of bibliographic elements for a variety of purposes, particularly for library applications.
-

Data Value Authorities

- **Getty Research Institute - Union Lists & Thesauri**
 - Thesaurus of Graphic Names (TGM) – <http://www.getty.edu/research/tools/vocabulary/tgn/>
 - Union List of Artist Names (ULAN) – <http://www.getty.edu/research/tools/vocabulary/ulan/>
 - Art and Architecture Thesaurus (AAT) – <http://www.getty.edu/research/tools/vocabulary/aat/index.html>
 - **Library of Congress - Authority Files & Thesauri**
 - Subject Headings (LCSH) – <http://authorities.loc.gov/>
 - Thesaurus for Graphic Materials: Subject Terms (TGMI) – <http://www.loc.gov/lexico/servlet/lexico/>
 - TGM: Genre and Physical Characteristic Terms (TGMII) – <http://www.loc.gov/lexico/servlet/lexico/>
-

Data Exchange / Transmission Standards

- **Metadata Encoding Transmission Standard (METS)** – <http://www.loc.gov/standards/mets>
METS is an XML schema for encoding structural metadata about complex digital objects and acts as a container with places to insert descriptive, administrative, and technical metadata.
- **Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH)** – <http://www.openarchives.org/>
Mechanism for repository interoperability wherein repositories, or *Data Providers*, expose

APPENDIX F: Standards, Best Practices, and Tools

structured metadata via OAI-PMH to *Service Providers* that in turn make OAI-PMH service requests to harvest that metadata.

- **XML Schema for the Sitemaps & Sitemap Index Files**

The following XML schemas define the elements and attributes that can appear in your Sitemap file. You can download this schema from the links below:

- **For Sitemaps:** <http://www.sitemaps.org/schemas/sitemap/0.9/sitemap.xsd>
 - **For Sitemap Indexes:** <http://www.sitemaps.org/schemas/sitemap/0.9/siteindex.xsd>
-

Part 2: BEST PRACTICES

Best Practices Identified for Use Include:

Encoded Archival Description (EAD) Best Practices / Encoding Guidelines

- EAD Application Guidelines, version 1.0
- RLG Best Practice Guidelines for EAD
- OAC Best Practice Guidelines for EAD

Dublin Core (DC) Best Practices / Encoding Guidelines

- Dublin Core Library Application Profile
- CDP Dublin Core Metadata Best Practices

Bibliographic Best Practices / Encoding Guidelines

- MODS Implementation Guidelines for Cultural Heritage Materials

Encoded Archival Description (EAD) Best Practices / Encoding Guidelines

EAD Application Guidelines, version 1.0 – <http://www.loc.gov/ead/ag/aghome.html>

These official guidelines published by the Society of American Archivists (SAA) and Library of Congress, give general advice on how to use all elements in the EAD Tag Library.

RLG Best Practice Guidelines for EAD – www.oclc.org/programs/ourwork/past/ead/bpg.pdf

EAD best practice based to enhance interoperability of EAD-encoded finding aids both within and across institutions that is based on SAA's *EAD Application Guidelines, version 1.0*, but more specific.

OAC Best Practice Guidelines for EAD – <http://www.cdlib.org/inside/diglib/guidelines/bpgead/>

A core set of practices for encoding a subset of EAD required for use in finding aids contributed to the Online Archive of California that draws from the RLG & SAA EAD Best Practice Guidelines, but is more specific in scope.

Dublin Core (DC) Best Practices / Encoding Guidelines

Dublin Core Library Application Profile – <http://dublincore.org/documents/library-application-profile/>

Profile clarifies the use of the DC for libraries and related applications & projects by outlining required or recommended use of elements, qualifiers, controlled vocabularies, etc. drawn from other metadata schema.

CDP Dublin Core Metadata Best Practices –

<http://www.cdpheritage.org/cdp/documents/CDPDCMBP.pdf>

These guidelines for creating metadata for cultural heritage resources use simplified Dublin Core that targets a variety of formats and institutions, but focuses mainly on descriptive metadata for resource discovery.

Bibliographic Best Practices / Encoding Guidelines

APPENDIX F: Standards, Best Practices, and Tools

MODS Implementation Guidelines for Cultural Heritage Materials –

http://www.diglib.org/aquifer/DLF_MODS_ImpGuidelines_ver4.pdf

Guidelines on the use of the MODS metadata schema to provide shareable metadata describing cultural materials, including books, images, film and physical artifacts. The metadata is mostly descriptive with some administrative elements, but structural & preservation metadata are not included.

Part 3: TECHNOLOGY TOOLS

Tools Identified for Use Include:

SGML/XML Encoding Tools

- EAD Cookbook
- OAC EAD Web Templates
- EAD Toolkit

Conversion / Transformation Tools

- EAD to MARC Conversion Script
- OAC EAD Conversion Tool
- SAA XSL Stylesheets
- LOC Conversion Toolkit

XML Validation Tools

- EAD Validation
- Sitemap Validation

Metadata Crosswalk Tools

- Library of Congress, MARC Standards: MARC21 Formats Mappings
- Getty Research Institute, Metadata Standards Crosswalks:

Unique ID Tools

- Archival Resource Key (ARK)
- MARC Organization Code

Open Source Archival Management Systems

- Archon
- Archivists' Toolkit (AT)
- International Council on Archives-Access to Memory (ICA-AtOM)

SGML/XML Encoding Tools

- **EAD Cookbook** – <http://www.archivists.org/saagroups/ead/ead2002cookbookhelp.html>
Provides authoring tools for XML Editors, such as Oxygen or XMetal, that make it easier to create EAD Finding Aids by providing a set of stylesheets used to transform the XML into HTML, and guidance about how to publish Finding Aids online, plus step-by-step assistance with implementation of EAD.
- **OAC EAD Web Templates** – <http://www.cdlib.org/inside/projects/oac/toolkit/templates/>
These online forms were developed by the OAC for generating collection- through series-/subseries-level finding aids that are compliant with the OAC BPG EAD and EAD Version 2002. Encoders cut and paste segments of their non-EAD finding aids into the form. The form is then converted to a text file and saved as a XML EAD file.
- **EAD Toolkit** – <http://www.cdlib.org/inside/projects/oac/toolkit/>
This "toolkit" of freeware and OAC- and MOAC-developed resources is provided for contributing members to assist in the creation, editing, manipulation, proofing, and viewing of EAD Version

APPENDIX F: Standards, Best Practices, and Tools

2002 finding aids that are compliant with the [OAC Best Practice Guidelines for Encoded Archival Description](#).

Conversion / Transformation Tools

- **EAD to MARC**
 - **EAD to MARC Conversion Script** <http://staff.washington.edu/carlsonm/>
Stylesheets, scripts, and support files for converting EAD finding aids into MARC21.
 - **EAD Version 1.0 to EAD Version 2002**
 - **OAC EAD Conversion Tool** <http://texts-stage.cdlib.org/xtf/preview.html>
Simple web form to convert finding aids by adding a valid URL for the XML finding aid in the text field to triggers the SAA conversion XSL stylesheets so users don't have to implement them locally.
 - **SAA XSL Stylesheets** – <http://www.archivists.org/saagroups/ead/resources/ead2002conv/>
Contains links to a zip archive of XSL conversion stylesheets, documentation, and associated binary tools for converting EAD Version 1.0 finding aids to EAD Version 2002.
 - **LOC Conversion Toolkit** – http://lcweb2.loc.gov/music/eadmusic/eadconv12/ead2002_r.htm
Provides a simple and bare bones open source conversion toolkit for EAD 1.0 to EAD 2002.
-

XML Validation Tools

- **EAD Validation Tools**
 - **OAC BPG Validator** – <http://www.cdlib.org/inside/projects/oac/toolkit/validator/>
This desktop XML parser was developed by the OAC to provide users with a log of EAD Version 2002 encoding errors, as well as OAC Best Practices Guidelines EAD compliance errors, with every error linked to a specific line in the EAD file so users can click on and easily view each.
 - **EAD “Report Card”** – <http://www.rlg.org/ead-report-card/>
Automated web application for checking the quality of your EAD encoding that supplements the RLG Best Practice Guidelines for Encoded Archival Description.
 - **Sitemap Validation Tools**

There are a number of tools available to help you validate the structure of your Sitemap based on this schema. You can find a list of XML-related tools at each of the following locations:

 - Official W3 Sitemap Validation Tool – <http://www.w3.org/XML/Schema#Tools>
 - Official XML Sitemap Validation Tool – <http://www.xml.com/pub/a/2000/12/13/schematools.html>
-

Metadata Crosswalk Tools

- **Getty Research Institute, Metadata Standards Crosswalks** – http://www.getty.edu/research/conducting_research/standards/intrometadata/crosswalks.html.

APPENDIX F: Standards, Best Practices, and Tools

Crosswalks relevant to art, architecture, & cultural heritage information maintained by Getty Standards.

- **Library of Congress, *MARC Standards: MARC21 Formats Mappings*** – <http://www.loc.gov/marc/marcdocz.html>. Crosswalks to and from MARC 21.
-

Unique ID Tools

- **MARC Organization Code** – <http://www.loc.gov/marc/organizations/>
Any organization may request a code from the Library of Congress to identify the collecting repository or institution in the <eadid> element using the MAINAGENCYCODE attribute value.
 - **Archival Resource Key (ARK)** – <http://www.cdlib.org/inside/diglib/ark/>
URL created to allow persistent, long-term access to information objects and can identify objects of any type: digital documents, databases, images, software, and websites, as well as physical objects (bones, statues, etc.) or intangible objects (chemicals, etc). Any institution may obtain a NAAN code to assign ARKs which are captured in the <eadid> element using the IDENTIFIER attribute value.
-

Open Source Archival Management Systems

Archival management systems may be less flexible than EAD creation tools, and getting legacy data into these systems can be challenging. However, they offer a number of features that may lead to greater efficiency and sustainability, such as support for authority control, reduced redundancy of data, easy data entry interfaces, the ability to analyze archival data through the generation of reports, and Web-publishing capabilities.

1. **Archon** (<http://www.archon.org>)
Developed by archivists at the University of Illinois at Urbana-Champaign, Archon makes it easy for archives to publish their finding aids online by automating many technical tasks, such as producing an EAD or MARC record. In addition, staff members do not need to learn technical coding and need little or no training to use the system so long as one has a familiarity with basic principles of archival arrangement and description. Archon, which is built on PHP 5 and MySQL, enables archivists to capture information about accessions, create and publish finding aids online, and export EAD and MARC. A digital library module supports presenting digital objects along with finding aids. Archon is easy to customize and provides support for authority control. Others caution, however, that importing existing finding aids into Archon can be difficult, given the variability of EAD.
2. **Archivists' Toolkit (AT)** (<http://www.archiviststoolkit.org/>)
Developed by a consortium including the University of California, San Diego Libraries, the New York University Libraries, and the Five Colleges, Inc., Libraries and supported by The Andrew W. Mellon Foundation, AT bills itself as "the first open source archival data management system to provide broad, integrated support for the management of archives." AT uses a Java desktop client and a database back-end (MySQL, MS SQL, or Oracle). Users report that AT makes it easier to produce finding aids and export EAD and MARC, generates useful reports, provides robust authority control, and offers good support for standards such as METS. Archivists noted that it can be difficult to import existing finding aids and make AT accommodate existing workflows. AT does not yet provide Web-publishing capabilities.
3. **International Council on Archives-Access to Memory (ICA-AtOM)** (<http://www.ica-atom.org/>)

APPENDIX F: Standards, Best Practices, and Tools

ICA-AToM is open source, Web-based archival description software that aims to make it easy for archives to provide online access to their archival holdings, adhere to ICA standards, and support multiple collection types (even multirepository implementations) through flexible, customizable software. ICA-AToM is designed to support aggregation of data from multiple institutions through OAI, IETF Atom Publishing Protocol (APP), and possibly other mechanisms. ICA-AToM aims to distinguish itself through its support for translation and internationalization, basis in ICA standards such as ISAD-G and ISAD-H, flexibility and customizability, and ease of installation and use. The developers want ICA-AToM to become a platform to manage archival information, including creating digital repository interfaces to systems such as DSpace and Fedora through a plug-in architecture. They plan to build in Web 2.0 features such as user-contributed content, user tagging, and social networking. ICA-AToM is currently in beta testing. Version 1.2, due to be released in summer 2009, will provide support for accessioning, OAI harvesting, crosswalking to standards such as DACS, EAD import and export, and many other features.

PART 4: RESOURCES TO CONSIDER FOR FUTURE USE

Standards, Best Practices & Tools Under Consideration Include

Administrative Metadata – Preservation, & Rights

- PREMIS (Preservation Metadata: Implementation Strategies)
- Schema for Rights Declaration OAC Best Practice Guidelines for EAD

Miscellaneous Resources Identified for Future Consideration

- Digital Repositories (Local, Regional, National, International)
- Metadata Extraction Tools
- OTHERS WILL CONTINUE TO BE IDENTIFIED

NOTE: Not In Current Scope of Grant – Listed For Future Use Only

Administrative Metadata – Imaging, Preservation, & Rights

PREMIS (Preservation Metadata: Implementation Strategies) -

<http://www.oclc.org/research/projects/pmwg/>

PREMIS builds on the OAIS reference model and articulates a core set of preservation data defined as “ things most working preservation repositories are likely to need to know in order to support digital preservation” (1). The dictionary itself is based on a simple model which incorporates and refines concepts from the OAIS reference model. Five entities are identified:

- Intellectual Entity – concerned with descriptive metadata describing the overall entity being preserved
- Objects and Events – are concerned with the documentation of preservation metadata
- Agents – identifies agents associated with rights management
- Rights – records agreements with rights-holders that allow repository to take actions with an object

Schema for Rights Declaration - <http://www.loc.gov/standards/rights/METSRights.xsd>

Offered as part of the METS framework, this Rights Declaration schema has 3 main elements:

- Simple declaration of type of rights (copyright, licensed, etc.) and statement of that Rights Declaration
- The naming of the Rights Holder(s) with appropriate contact information
- The Context(s) for the rights declaration based on type of users who have a set of permissions for a digital object or part of a digital object or if there are any constraints to the permissions.

Open Source Digital Repositories

Rice University’s Digital Repository using DSpace – Rice University has developed a Digital Repository (RUDR) using DSpace, which is an open source digital asset management software platform that enables institutions to capture and describe digital content. It runs on a variety of hardware platforms and supports OAI-PMH version 2.0. Possibilities to use this repository in the future are currently being explored.

APPENDIX F: Standards, Best Practices, and Tools

Metadata Extraction Tools

NLNZ Metadata Extraction Tool <http://www.natlib.govt.nz/en/whatsnew/4initiatives.html#extraction> The National Library of New Zealand metadata extraction tool extracts metadata from various file types to generate preservation metadata.

Best Practices Selected to Guide Development of Project XML Models

- **EAD Models**
 - Library of Congress – Minimum Set of EAD Elements Required
 - Texas Archival Resources Online (TARO) –EAD Model Guidance
 - **XML Sitemap Protocol and Sitemap Indexes**
-

APPENDIX G: Selected XML Models

Library of Congress – Minimum Set of EAD Finding Aid Elements Required

Source: <http://www.loc.gov/ead/ag/agappa.html>

The EAD elements listed below comprise the minimum elements recommended for creation of a very basic EAD-encoded finding aid. This list includes both the elements required by the EAD DTD for validation (shown in **bold** type) and additional important structural elements, which together comprise "best practice" for a basic archival description. The list of elements assumes that the archival description begins at the collection, fonds, or series level, but many of the elements listed are applicable at any level of description.

As the brief list of **bolded** element names reveals, only a handful of elements are *required* to enable an EAD-encoded finding aid to be validated against the specifications of the DTD. This is in part due to reasons related to the SGML rules for writing DTDs, and also because the EAD developers recognized that many "legacy" finding aids do not contain a full or consistent set of data elements.

The final use & organization of EAD elements will be arranged to best meet the needs of the users, with the exception of the sections that have a *required* order as dictated by the EAD DTD, including the <eadheader> subelements, the high-level <did> within <archdesc>, and the <did> within <c> or <c0x>.

```
<ead>
  <eadheader>
    <eadid>
    <filedesc>
      <titlestmt>
        <titleproper>
        <author>
      <publicationstmt>
        <publisher>
        <date>
    <profiledesc>
      <creation>
      <language>
      <language>
  <archdesc> with LEVEL, LANGMATERIAL, and LEGALSTATUS attributes
    <did>
      <repository>
        <corpname>
      <origination>
        <persname>, <corpname>, <famname> as appropriate
      <unittitle>
      <unitdate>
      <physdesc>
      <unitid> with COUNTRYCODE and REPOSITORYCODE attributes
      <abstract>
    <admininfo>
      subelements as appropriate
    <bioghist>
    <scopecontent>
    <controlaccess>
      subelements as appropriate
    <dsc> with TYPE attribute
      <c0x> or <c> with LEVEL attribute in as many levels as appropriate
      <did>
        <container>
```

APPENDIX G: Selected XML Models

<unittitle>
other subelements as appropriate

Texas Archival Resources Online (TARO) –Model EAD Finding Aid & Guidance

Source: <http://www.lib.utexas.edu/taro/admin/howdoi/index.html>

```
<?xml version="1.0" encoding="UTF-8" standalone="no" ?>
<!DOCTYPE ead PUBLIC "-//ISBN 1-931666-00-8//DTD ead.dtd (Encoded Archival
  Description (EAD) Version 2002)//EN" "ead.dtd">
<ead relatedencoding="MARC21">
<eadheader audience="internal" langencoding="ISO 639-2"
  encodinganalog="local choice"
  <eadid countrycode="us" mainagencycode="\[ISO15511 code for repository\]>
  urn:taro:\[repository code\]:\[file name\]</eadid>
  <filedesc>
    <titlestmt>
      <titleproper>eg. James Smith -- used to display title of
        finding aid, and for fielded searching via a dc.title HTML metatag
      </titleproper> <subtitle>eg. An inventory of the
        James Smith Papers -- also used to display finding aid title
      </subtitle>
    </titlestmt>
  </filedesc>
</eadheader>

<archdesc>
  <did>
  <head></head>
  <repository>(with <corpname> and <subarea> subelements as appropriate)
    content used in repository fielded searching of the finding aids

    <extref href="http://repository.url.here/" />

  </repository>
  <origination>(with <corpname> and <persname> subelements as appropriate)
    content used in author fielded searching of the finding aids
  </origination>
  <unittitle>
    <unitdate type="inclusive">contents of each of these
      elements used in the OVERVIEW section of HTML view of the finding aid
    </unitdate>
  </unittitle>
  <abstract>as needed</abstract>
  <physdesc>would display in OVERVIEW section of HTML view of
    the finding aid </physdesc>
  <unitid> if used, countrycode and repositorycode attributes to be
    included; would display in OVERVIEW section of HTML view of the
```

APPENDIX G: Selected XML Models

```
finding aid </unitid>
      <physloc>would display in OVERVIEW section of HTML view of the
finding aid </physloc>
      <note>as needed</note>
    </did>
    <controlaccess>
      Use subelements as needed for each collection. </controlaccess>
    </archdesc>
  </lead>
```

TARO REPOSITORY MAIN AGENCY CODES

TARO website indicated Rice University Code = "txhr" <http://www.loc.gov/marc/organizations/org-search.php>

EAD Elements

<eadid> EAD Identifier

Description: A required subelement of <eadheader> that designates a unique code for a particular EAD finding aid document.

Two of the attributes, COUNTRYCODE and MAINAGENCYCODE, are required to make the <eadid> compliant with ISAD(G) element 3.1.1. MAINAGENCYCODE provides the ISO 15511 code for the institution that maintains the finding aid (which may not be the same as the institution that is the custodian of the materials described). COUNTRYCODE supplies the ISO 3166-1 code for the country of the maintenance agency. In addition to these two attributes, it is recommended that repositories also use at least one of the following attributes: URL, PUBLICID, or IDENTIFIER to make the <eadid> globally unique. PUBLICID should be a Formal Public Identifier, URL an absolute or relative address, and IDENTIFIER a machine-readable unique identifier for the finding aid file. (The proper syntax for PUBLICID is defined in ISO/IEC 9070:1991 Information technology -- SGML support facilities -- Registration procedures for public text owner identifiers.)

May occur within: eadheader

Examples:

```
<eadid countrycode="us" mainagencycode="txu-hu"
  publicid="-//us::txu-hu//TEXT us::txu-hu::hrc.00001//EN"
  url="www.lib.utexas.edu/taro/hrc/00001.xml">
  hrc.00001
</eadid>
```

File naming and identification

EADID: Outsourced XML files will return with an empty <eadid> element nested within the <eadheader>. EADID requires a formatted string for its content prior to sending the file to the TARO production server.* The content format remains the same whether you are tagging to the EAD 1.0 or EAD 2002 specification.

FOR: Rice, Texas A&M, Texas Tech, University of Houston and UT Austin repositories.

APPENDIX G: Selected XML Models

The content for <eadid> needs to follow this format: urn:taro:[*institution*].[*unit*].[*xxxxx*]

- **where. . .**
- *institution* = uh, utexas, rice, tamu, ttu
- *unit*
 - for texas = hrc, cah, law, aaa, blac
 - for tamu = cush
 - for ttu = sw, swcpc, rb, ua
 - for rice = wrc
 - for uh = ua, sc, warc, me
- *xxxxx* = a unique FIVE DIGIT number, starting with 00001, assigned to each finding aid

File names: The name you give each XML file must coincide with the last five digits of the URN in the <eadid> of that file. In this example, if the urn is: urn:taro:utexas.hrc.00012 then you would name the XML file -- 00012.xml

EAD 2002 also [requires the presence of certain attributes for <eadid>](#).

<DID> should not be considered optional

"Imagine a typical scenario: An archivist begins encoding a finding aid by first opening the EAD element and creating the required EADHEADER. He or she may add some FRONTMATTER before opening the ARCHDESC element and setting its LEVEL attribute to the value "collection," "record group," "fonds," or "series," depending on which term best reflects the character of the whole unit being described in the finding aid. What then follows are data elements that describe that whole unit, including a special subset of core data elements that are gathered together under a parent element called Descriptive Identification (DID). *These DID subelements are thought to be among the most important for ensuring a good basic description of an archival unit or component.* Grouping these elements together serves several purposes. It ensures that the same data elements and structure are available at every level of description within the EAD hierarchy, facilitates the retrieval or other output of a cohesive body of elements for resource discovery and recognition, and, because the elements appear together in the tag library and on software menus and templates, helps to remind encoders to capture descriptive information they may otherwise overlook.

EAD Tag Library Version 1.0

*For linking to other pages or sites use <extref> with the attribute href="..."
(for example: use with links in repository)*

Extended Reference

Description: A linking element that can include text and subelements as part of its reference to an electronic object that is external to the EAD document. For TARO, use the HREF attribute to identify the external object.

EXAMPLE:

```
<repository label="Repository" encodinganalog="852$a">  
  <extref href="http://www.cah.utexas.edu/" show="new" actuate="onrequest">  
    <corpname encodinganalog="852$a">  
      <subarea>Center for American History,</subarea>  
      The University of Texas at Austin
```

APPENDIX G: Selected XML Models

```
</corpname>
</extref>
<repository>
```

REPOSITORY subelements

As mentioned elsewhere, the REPOSITORY tag (as found within the DID) is being used to generate the dc.publisher metatag in the HTML version of the finding aids. A search of this metatag is available within the form on the TARO finding aids website. It is also used in generating the website browse pages for each repository. In other words, the presence of this tag in each find aid provides an important point of access.

ORIGINATION subelements

PERSNAME: *last name, first name middle name*

CORPNAME: Titles of works MUST drop the initial article.

- bad: The New York Times
- good: New York Times

XML Sitemap Protocol and Sitemap Indexes

SOURCE: <http://sitemaps.org/>

The Sitemap protocol format consists of XML tags that list URLs for a site along with additional metadata about each URL (last updated, how often it changes, and how important it is, relative to other URLs in the site) so that search engines can more intelligently crawl the site. The Sitemap protocol also enables users to provide details about web pages to search engines in order to provide additional information access points beyond just the URLs.

Sample XML Sitemap Index

The following example shows a Sitemap index that lists two Sitemaps:

```
<?xml version="1.0" encoding="UTF-8"?>
<sitemapindex xmlns="http://www.sitemaps.org/schemas/sitemap/0.9">
  <sitemap>
    <loc>http://www.example.com/sitemap1.xml.gz</loc>
    <lastmod>2004-10-01T18:23:17+00:00</lastmod>
  </sitemap>
  <sitemap>
    <loc>http://www.example.com/sitemap2.xml.gz</loc>
    <lastmod>2005-01-01</lastmod>
  </sitemap>
</sitemapindex>
```

Sample XML Sitemap

The following example shows a Sitemap in XML format. The Sitemap in the example contains a small number of URLs, each using a different set of optional parameters.

```
<?xml version="1.0" encoding="UTF-8"?>
<urlset xmlns="http://www.sitemaps.org/schemas/sitemap/0.9">
  <url>
    <loc>http://www.example.com/</loc>
    <lastmod>2005-01-01</lastmod>
    <changefreq>monthly</changefreq>
```

APPENDIX G: Selected XML Models

```
<priority>0.8</priority>
</url>
<url>
  <loc>http://www.example.com/catalog?item=12&amp;desc=vacation_hawaii</loc>
  <changefreq>weekly</changefreq>
</url>
<url>
  <loc>http://www.example.com/catalog?item=73&amp;desc=vacation_new_zealand</loc>
  <lastmod>2004-12-23</lastmod>
  <changefreq>weekly</changefreq>
</url>
<url>
  <loc>http://www.example.com/catalog?item=74&amp;desc=vacation_newfoundland</loc>
  <lastmod>2004-12-23T18:00:15+00:00</lastmod>
  <priority>0.3</priority>
</url>
<url>
  <loc>http://www.example.com/catalog?item=83&amp;desc=vacation_usa</loc>
  <lastmod>2004-11-23</lastmod>
</url>
</urlset>
```

Below is specific guidance provided by the sitemaps.org group for creating sitemaps.

In order to be valid the Sitemap file must:

- Begin with an opening `<urlset>` tag and end with a closing `</urlset>` tag,
- Specify the namespace (protocol standard) within the `<urlset>` tag,
- Include a `<url>` entry for each URL, as a parent XML tag,
- Include a `<loc>` child entry for each `<url>` parent tag, and
- All URLs in a Sitemap must be from a single host.

Entity escaping

Your Sitemap file must be UTF-8 encoded (you can generally do this when you save the file). As with all XML files, any data values (including URLs) must use entity escape codes for the characters listed in the table below.

Character		Escape Code
Ampersand	&	&
Single Quote	'	'
Double Quote	"	"
Greater Than	>	>
Less Than	<	<

Using Sitemap index files (to group multiple sitemap files): If you do provide multiple Sitemaps, you should then list each Sitemap file in a Sitemap index file. Sitemap index files may not list more than 50,000 Sitemaps and must be no larger than 10MB (10,485,760 bytes) and can be compressed. You can have more than one Sitemap index file. The XML format of a Sitemap index file is very similar to the XML format of a Sitemap file.

The Sitemap index file must:

APPENDIX G: Selected XML Models

- Begin with an opening `<sitemapindex>` tag and end with a closing `</sitemapindex>` tag.
- Include a `<sitemap>` entry for each Sitemap as a parent XML tag.
- Include a `<loc>` child entry for each `<sitemap>` parent tag.

The optional `<lastmod>` tag is also available for Sitemap index files.

Validating your Sitemap: There are a number of tools available to help you validate the structure of your Sitemap based on this schema. You can find a list of XML-related tools at each of the following locations:

- <http://www.w3.org/XML/Schema#Tools>
- <http://www.xml.com/pub/a/2000/12/13/schematools.html>

Informing search engine crawlers: Once the Sitemap file is created and placed on the webserver, the next step is to inform the search engines that support this protocol of its location. The search engines can then retrieve the Sitemap and make the URLs available to crawlers. This can be done by:

- Submitting it to them via the search engine's submission interface
- Specifying the location in your site's robots.txt file
- Sending an HTTP request

IMPORTANT NOTE: *The decision regarding this project's use of the **MARC21 / MARCXML** and/or **MODS** standards will be made finalized closer to the start of project in order to reassess the most recent recommendations from the Archival community at that time or new conversion tools that have been made available between the now and then.*