

Texas Heritage Digitization Initiative

Standards for Cultural Heritage Digitization Projects

Approved by:

THDI Steering Committee

February 22, 2007

General Introduction

Hundreds of institutions in Texas hold significant collections that document cultural heritage. Libraries, museums, archives, historical societies and governmental agencies have lovingly collected and cared for manuscripts, photographs, maps, publications and many other items that are important resources for students, scholars, policy makers, lifelong learners and many other audiences. The Texas Heritage Digitization Initiative (THDI) envisions unified online access to these cultural heritage resources held by Texas institutions to enhance understanding of our cultural heritage and better serve these users.

In May of 2004, representatives of libraries, museums, archives and governmental agencies responded to an invitation by the Texas State Library and Archives Commission to explore this vision and propose a plan to achieve it. More than 80 individuals attended this initial meeting, with a majority volunteering their time to serve on one of six Work Groups. The initial Standards Work Group, chaired by Brian Surratt of Texas A&M University, developed a framework for thinking about standards in the context of practices used by Texas cultural heritage institutions. In 2006, the Standards Work Group continued that process and developed the basic standards in this document.

The Texas Heritage Digitization Initiative's *Standards for Cultural Heritage Digitization Projects* are not the first or by any means the only attempt to set out guidelines for digitization projects. However, they take a slightly different approach. These Standards are in reality a collection of standards in various areas, including digitization, metadata, controlled vocabulary, interoperability, and preservation management. The available standards in each area are divided into three categories: minimal, which is the lowest level of common practice that the Work Group found to be acceptable; basic, which is the level that most projects should attempt to meet; and enhanced, which is generally intended for researchers and others who have an ongoing commitment to continually updating their digital collections to use the newest tools and standards available.

Each individual project may decide for itself which category to use in any given area. Often, cost is a significant factor in this decision. For example, institutions may have a Digital Asset Management System that supports the Dublin Core metadata syntax but not newer syntaxes such as MODS, and upgrading the system would require significant investments in software and staff development. There is nothing wrong with an institution deciding that, for its purposes, Dublin Core is perfectly adequate. There is also nothing wrong with an institution deciding to make an investment in more expensive, but possibly more descriptive, tools for describing and storing its digital collections.

In 2007, the Standards Work Group will turn its attention to preparing a set of Best Practices for *Cultural Heritage Digitization Projects*, which will provide more practical advice for digital project managers on issues such as organizing digital objects, assigning titles and dates to both physical objects and their digital surrogates, and other topics. We invite your comments and suggestions as we work through this process. This will not be the final word on standards and best practices. The THDI Standards Work Group expects to update these documents regularly in order to incorporate new standards, tools, and techniques as they are developed. To follow our progress, check the area for the Standards Work Group at www.thdi.org.

Jerry Drake, General Land Office
Chair, THDI Standards Work Group, 2006-2007

Houston McGaugh, Star of the Republic Museum
Chair, THDI Steering Committee, 2006-2007

Danielle Cunniff Plumer, THDI Coordinator
Texas State Library and Archives Commission

Digitization Standards

Introduction

Presenting cultural heritage materials online requires conversion, whether from an original physical object to a digital surrogate, an analogue surrogate (such as microfilm) to a digital surrogate, or from one digital format to another. Creating digital surrogates from two-dimensional still images and texts is a relatively well-understood process, and several guides to best practices are available. Creating digital surrogates of three-dimensional objects and time-based media such as audio and video is less well understood. Participants should understand that this is an evolving field and that recommended practices may change over time, sometimes with the result that re-digitization will be required. In general, open standards are to be preferred over proprietary standards, as this may reduce the need for re-digitization in the future.

Excellent overviews of the digitization process can be obtained from the Cornell University Library's *Moving Theory Into Practice: Digital Imaging Tutorial* and from the Getty Institute's *Introduction to Imaging (Revised Edition)* by Howard Bessner, edited by Sally Hubbard with Deborah Lenert.

Levels of Digitization:

THDI recognizes three levels of digitization:

Minimal

Participants will provide online access to "reference" or access-quality digital objects, objects that are designed for use in the context of the World Wide Web. These objects may be images, texts, or multimedia documents, and may appear in any of the formats commonly supported by Web browsers. For specific formats and resolutions, refer to the *Western States Digital Imaging Best Practices, Version 1.0*, and *CDP Digital Audio Best Practices, Version 2.0*.

Some older collections may include materials that do not meet current standards, possibly because they were scanned using older standards or using equipment that was not capable of capturing sufficiently high resolutions or saving to uncompressed formats. Ideally, these objects would be re-digitized, but until such time as that is possible, they should be treated as meeting minimal standards.

Basic

Participants will create "master" or archival quality digital objects, objects that are saved in a lossless compression format at a sufficiently high resolution or sampling rate that the object is a "reasonable reproduction" of the original, without enhancement, as described in the 2004 *NARA Technical Guidelines for Digitizing Archival Materials for Electronic Access*. See the "Recommended Image Parameters" in the NARA Guidelines for specific minimum resolutions for text and still images. Technical metadata, including at a minimum EXIF data if generated by a digital still camera, will be created for each object.

In addition, participants will provide online access to "reference" or access-quality digital objects and to "thumbnails" of those access-quality digital objects. In the case of multimedia objects, thumbnails may be extracted from the original sequence or a "snippet" of the file may be provided for reference, and highly-compressed versions of the master object may be made available for streaming or download.

Enhanced

Participants will create "master" or "archival" quality digital objects that meet the requirements for a level 2 image quality assessment rating according to the *NARA Technical Guidelines for Digitizing Archival Materials for Electronic Access*. Detailed technical metadata about each object will be generated according to the NISO Z39.87 draft standard or its successor(s) and will be stored in both the image header, where possible, and in an external database. Participants may choose to create access-quality

digital objects and "thumbnails" or "snippets" from the master object or may enable automated creation of such objects through digital library systems.

References

- Bessner, Howard, edited by Sally Hubbard with Deborah Lenert. 2003. *Introduction to Imaging*. Revised Edition. Retrieved 2 February 2006 from the Getty Institute's website:
http://www.getty.edu/research/conducting_research/standards/introimages/
- CDP Digital Audio Working Group. November 2005. *Digital Audio Best Practices*, Version 2.0. Retrieved 12 January 2006 from http://www.cdpheritage.org/digital/audio/documents/CDPDABP_1-2.pdf
- Cornell University Library. 2003. *Moving Theory Into Practice: Digital Imaging Tutorial*. Retrieved 2 February 2006 from <http://www.library.cornell.edu/preservation/tutorial/>.
- National Information Standards Organization, and AIIM International. 2005. *Data Dictionary: Technical Metadata for Digital Still Images* (NISO Z39.87-200X; AIIM 20-200X). Draft Standard. Retrieved 12 January 2006 from <http://www.niso.org/standards/resources/Z39-87-200x-forballot.pdf>
- Puglia, Steven, Jeffrey Reed, and Erin Rhodes. June 2004. *U.S. National Archives and Records Administration (NARA) Technical Guidelines for Digitizing Archival Materials for Electronic Access: Creation of Production Master Files – Raster Images For the Following Record Types- Textual, Graphic Illustrations/Artwork/Originals, Maps, Plans, Oversized, Photographs, Aerial Photographs, and Objects/Artifacts*. Retrieved 12 January 2006 from
<http://www.archives.gov/research/arc/preservation/digitizing-archival-materials.pdf>
- Western States Digital Standards Group, Digital Images Working Group. 2003. *Western States Digital Imaging Best Practices*, Version 1.0. Retrieved 12 January 2006 from
http://www.cdpheritage.org/digital/scanning/documents/WSDIBP_v1.pdf

Metadata Standards

Introduction

Metadata consists of textual information about digitized resources. Metadata is most closely associated with bibliographic description in support of the search, retrieval, and identification of resources, including both physical and digital objects. Excellent resources for metadata come from the Getty Institute's *Introduction to Metadata: Pathways to Digital Information*, by Tony Gill, Anne J. Gilliland, and Mary S. Woodley, edited by Murtha Baca, and the NISO publication *Understanding Metadata*.

Metadata is actually composed of three elements. The first is a content standard, which sets out rules or guidelines for cataloging; AACR2, is one such standard, most frequently used in libraries. *Cataloging Cultural Objects* (CCO) is a new standard for museums, and *Describing Archives: A Content Standard* (DA:CS), is a standard for archival description. The second element is a syntax, which establishes out the specific metadata elements that are available; Dublin Core, MODS, and MARC are all syntaxes. The third element is format, which may include XML and ASCII. Format is important because specific software may be required to read a given metadata file.

Because metadata is often specific to the type of original resource being described, there is no one standard or syntax that best describes every type of resource. Crosswalks have been developed to map metadata from one syntax to another; one such crosswalk is described in the Getty Institute's *Introduction to Metadata*. Participants are encouraged to identify the type of metadata best suited to their collections and to describe the objects in their collections as fully as possible using whatever metadata syntax they have selected.

Metadata can be divided into various types: descriptive, used to provide information about the content, subject, or composition of the object, particularly in order to provide resource identification and discovery; structural, used to describe how parts of a complex object relate to each other; and administrative, such as that recorded as part of the digital object's lifecycle of creation, acquisition, use, preservation, and (perhaps ultimately) deletion. Some metadata specialists add additional types, such as preservation metadata (see the section in these standards on Preservation) or technical metadata, which is often created automatically by the device used to create a digital object.

Metadata should be developed regardless of the search or browse mechanism planned to provide access to the digital objects in a collection. However, participants should be aware that certain types of metadata provide more search functionality than others and that the digital asset management system selected to store the metadata (and possibly the digital objects themselves) will also expand or restrict search functionality. For more information on search functionality, participants should read the "Interoperability" section of this document.

Levels of Metadata

THDI recognizes three levels of metadata:

Minimal

Participants will provide access to metadata about digital objects in their collection. The simplest form of metadata consists of simple text, sometimes in the form of "keywords," or terms chosen from an uncontrolled vocabulary to describe the resource. "Tags" and captions may also be considered metadata. This form of metadata may be visible to the user or may be embedded in an HTML or other file.

To be searchable, metadata of this type must be indexed, or "spidered," as for example is done by Google™ and other search engines. Participants are strongly encouraged to allow this type of indexing, although THDI at present has no plans to implement an indexing strategy. For this reason, collections meeting only minimal metadata standards will by default be excluded from the THDI search portal.

Basic

Participants will provide descriptive metadata for the digital objects in their collection at a sufficient level of granularity to distinguish individual objects using a metadata standard appropriate to their collection type; often, the choice of metadata may also be driven by the choice of a digital asset management system such as a library catalog or database. Some examples of appropriate metadata syntaxes include MARC, Dublin Core (including variants such as CDP, Western States, and UNTL, or more generally the DC-Library Application Profile), TEI or EAD headers, and the Content Standard for Digital Geospatial Metadata (CSDGM), but this list is not meant to be exclusionary.

Enhanced

In addition to descriptive metadata, participants will provide administrative metadata for all of the digital objects in their collection. Structural, technical, and preservation metadata should also be included whenever possible. Certain metadata syntaxes, particularly METS but also to a lesser extent MODS, Qualified Dublin Core, TEI, and EAD, allow the provision of these additional metadata types. Participants should regularly maintain and update their metadata as new guidelines and standards are established.

References

- Baca, Murtha (ed.), Patricia Harpring, Elisa Lanzi, Linda McRae, and Ann Whiteside. 2006. *Cataloging Cultural Objects: A Guide to Describing Cultural Works and Their Images*. Chicago : American Library Association, 2006
- Baca, Murtha (ed.), Tony Gill, Anne J. Gilliland, and Mary S. Woodley. 2000. *Introduction to Metadata: Pathways to Digital Information*. Online Edition, Version 2.1. Available at http://www.getty.edu/research/conducting_research/standards/intrometadata/
- CDP Metadata Working Group. 2005 September. *Dublin Core Metadata Best Practices*, Version 2.1. Available at from <http://www.cdpheritage.org/cdp/documents/CDPDCMBP.pdf>
- Dublin Core Metadata Initiative, National Information Standards Organization, and American National Standards Institute. 2001. *The Dublin Core Metadata Element Set*. Available at <http://www.niso.org/standards/resources/Z39-85.pdf>
- DCMI Usage Board. 2004. *DCMI Metadata Terms*. Available at from <http://www.dublincore.org/documents/dcmi-terms/>
- DCMI-Libraries Working Group. 2004. *DC-Library Application Profile* (DC-Lib). Available at <http://dublincore.org/documents/library-application-profile/>
- Federal Geographic Data Committee. 1998. *Content Standard for Digital Geospatial Metadata*. Version 2.0. Available at from <http://www.fgdc.gov/standards/projects/FGDC-standards-projects/metadata/base-metadata/>
- Library of Congress. 2005. *Encoded Archival Description (EAD): Official EAD 2002 Web Site*. Available at from <http://www.loc.gov/ead/>
- Library of Congress. 1999. *MARC 21 Format for Bibliographic Data: Including Guidelines for Content Designation*. 2 vols. Washington, D.C.: Library of Congress, Cataloging Distribution Service.
- Library of Congress. 2004. *METS: Metadata Encoding & Transmission Standard Official Web Site*. Available at from <http://www.loc.gov/standards/mets/>
- Library of Congress. 2004. *MODS User Guidelines* (Ver. 3.0). Available at from <http://www.loc.gov/standards/mods/v3/mods-userguide.html>
- NISO. 2004. *Understanding Metadata*. Bethesda, MD: NISO Press. Available at from <http://www.niso.org/standards/resources/UnderstandingMetadata.pdf>

OCLC/RLG Working Group on Preservation Metadata. 2002. *Preservation OAIS Information Model: A Metadata Framework to Support the Preservation Objects*. Available at http://www.oclc.org/research/projects/pmwg/pm_framework.pdf

Society of American Archivists. 2004. *Describing Archives: A Content Standard*. Chicago : Society of American Archivists.

Text Encoding Initiative Consortium. 2004. *TEI P4: Guidelines for Electronic Text Encoding and Interchange*, XML-compatible edition. Available at from <http://www.tei-c.org/P4X/>

University of North Texas Libraries. 2005 November. *The UNTL Metadata Guidelines*. Available at <http://www.library.unt.edu/digitalprojects/metadata/UNTL-Metadata-Guide.pdf>

Controlled Vocabulary Standards

Introduction

Controlled vocabularies allow catalogers and collection managers to use descriptive terminology consistently. Controlled vocabularies may take the form of **authority lists**, in which one form is considered to be definitive, **thesauri**, which describe relationships among various terms, and **structured classifications**, which generally create hierarchical and/or faceted relationships between terms. Controlled vocabularies are primarily used for subject headings, but they can also establish the "correct" spellings of individual terms and the preferred structure of complex terms, such as the formation of personal names as "last name, first name" or dates as "two-digit-date three-letter-month four-digit-year."

Properly developed controlled vocabulary supports both the search and browse functions of information retrieval, as described in Marcia J. Bates' article "[The Design of Browsing and Berrypicking Techniques for the Online Search Interface](#)." A controlled vocabulary may be used as a browse list directly, or a search result could return an item with a subject heading from a controlled vocabulary source that is a hyperlink enabling a search for items that also are described by that subject heading. The advantage of using controlled vocabulary instead of uncontrolled keywords is that this hyperlinking can be managed without the need for additional data mining or processing. While uncontrolled keywords remain useful tools for description (and are particularly useful in the new area of folksonomies), they should be considered a supplement to controlled vocabularies rather than a replacement.

The Getty Museum's [Introduction to Vocabularies: A Guide to Enhancing Access to Art and Material Culture Information](#) provides a good overview of controlled vocabulary, although it is scheduled to be updated, and participants are urged to consult the latest version when it becomes available.

Controlled Vocabulary Best Practices

The preferred THDI authority list for local and regional terms is [The Handbook of Texas Online](#), which includes articles on people, places, and events related to Texas history and heritage. The authoritative version of any given term is that used in the article title, which may or may not be supplemented at the local level with a date facet. The *Handbook of Texas Online* is limited, however, in that it does not include articles for living people and may not have developed articles on people, places, and events of strictly regional interest. Participants should therefore supplement the *Handbook of Texas Online* with other sources of controlled vocabulary.

Levels of Controlled Vocabulary

THDI recognizes three levels of controlled vocabulary use:

Minimal

Participants will use a locally developed and maintained controlled vocabulary for names and subject headings. This is particularly appropriate for regional terms, but participants should be aware that use of a non-standard controlled vocabulary may result in omission of some records from search results.

Basic

Participants will use one or more standard sources for controlled vocabulary, such as the Getty Vocabularies, Library of Congress Subject Headings and Authorities, Chenhall's Nomenclature, etc. Whenever possible, the source of the controlled vocabulary will be indicated by a namespace or code in the descriptive metadata record.

Some common sources of controlled vocabulary include the following:

- Alexandria Digital Library Gazetteer Server
<http://www.alexandria.ucsb.edu/clients/gazetteer/>
- Alexandria Digital Library Geographic Feature Type Thesaurus
<http://www.alexandria.ucsb.edu/gazetteer/FeatureTypes/ver070302/>
- Art and Architecture Thesaurus (AAT)
http://www.getty.edu/research/conducting_research/vocabularies/aat/
- Getty Thesaurus of Geographic Names (TGN)
http://www.getty.edu/research/conducting_research/vocabularies/tgn/
- Geographic Names Information System (GNIS)
<http://geonames.usgs.gov/>
- Handbook of Texas Online
<http://www.tsha.utexas.edu/handbook/online/>
- Library of Congress Authorities
<http://authorities.loc.gov/>
- Thesaurus for Graphic Materials I: Subject Terms
<http://lcweb.loc.gov/rr/print/tgm1/toc.html>
- Thesaurus for Graphic Materials II: Genre and Physical Characteristic Terms
<http://lcweb.loc.gov/rr/print/tgm2/>
- Union List of Artist Names (ULAN)
http://www.getty.edu/research/conducting_research/vocabularies/ulan/

Enhanced

Participants will participate in the development of controlled vocabulary appropriate to their projects, regions, and resources. For the *Handbook of Texas Online*, participants may use the "suggest an article topic" link if an article is not available for any given term, or the "report an error or correction" link for articles that are incorrect. Additional tools for controlled vocabulary development may be available for the *Handbook of Texas Online* in the future. The Getty Vocabularies also have a formal method for contributions from participants.

Participants may also wish to explore alternate methods for vocabulary development, including new applications for controlled vocabulary such as the OCLC [FAST project](#) or tools for generating uncontrolled vocabulary, sometimes referred to as "folksonomies" ([Wikipedia article](#)), or systems such as collabularies ([Wikipedia article](#)) that fall somewhere in between. However, folksonomies are primarily to be used as a supplement to controlled vocabularies rather than a replacement. Participants using locally developed vocabularies will post their local vocabularies online and establish a formal ontology that can be used in an xml namespace declaration.

References

- Lanzi, Elisa, and Patricia Harpring. (2000). Introduction to Vocabularies: A Guide to Enhancing Access to Art and Material Culture Information. Retrieved March 6, 2006, from http://www.getty.edu/research/conducting_research/vocabularies/introvocabs/
- Shirky, Clay. (2005). Ontology is Overrated: Categories, Links, and Tags. Retrieved March 13, 2006, from http://www.shirky.com/writings/ontology_overrated.html
- Texas State Historical Association. (2005). The Handbook of Texas Online. Available from <http://www.tsha.utexas.edu/handbook/online/>

Interoperability Standards

Introduction

Interoperability refers to the capacity of a system to interact with other systems in a predefined manner. THDI does not intend to create a single state-wide repository for all of Texas; rather, THDI will promote the development of institutional and regional collections in a variety of digital asset management systems using a variety of standards and best practices. However, to promote the best use of these repositories, and particularly to enable access to them through a single search interface, it will be necessary for THDI to specify particular interoperability requirements and standards.

Currently, there are three ways to provide a common search interface. The first is to develop software routines, commonly called "bots" or "spiders" to index Web-accessible materials. This index can then be searched and links provided to the original resources. This approach is the one currently taken by most major Web search engines but is not part of the THDI search interface plans. The second approach is most common in the library community and involves the use of a common query language and a standard interoperability profile developed according to the Z39.50 protocol to search disparate resources in real-time. Texas has been a leader in the development of a profile for statewide resources and in the deployment of a search interface to access library catalogs throughout the state. The third approach is the newest and involves the use a metadata harvester that accesses pre-identified collections known to have been developed according to the Open Archives Initiative's *Protocol for Metadata Harvesting*. A number of harvesters are in use nationally, but none has been developed in Texas at this time, although one of the results of THDI's IMLS grant will be an OAI harvester developed and run by the University of North Texas, which can then be searched via Z39.50 from the main THDI search interface.

Levels of Interoperability

THDI recognizes three levels of interoperability:

Minimal

Participants will make metadata available in the form of a static repository as defined by the Open Archives Initiative. Participants will create metadata in the oai_dc format and will create a single static repository XMP document which will be registered with a static repository gateway as described in the DLF/NSDL *Best Practices for OAI Static Repositories*.

Basic

Participants will make metadata available to be searched using either the Z39.50 protocol or its successors or derivatives, including SRU/SRW, or the Open Archives Initiative's *Protocol for Metadata Harvesting*.

If participants choose to use the OAI-PMH, they should follow the *Guidelines for Repository Implementers* and the DLF/NSDL *Best Practices for OAI Data Provider Implementations*. Participants are encouraged to use the OAI's XML schema to hold branding information for collections as well in order to differentiate their collections in THDI search results.

If participants choose to provide a Z39.50 implementation, it will conform to the *Z Texas Profile Functional Area C*, level 0. This area provides a definition of a core set of searches (and the associated attributes and attribute combinations) required for basic search and retrieval mechanisms needed by library users when interacting with disparate electronic resources, duplicating the Bath Profile Functional Area C. The following elements are used in this area/level:

- Creator Search -- Keyword
- Title Search -- Keyword
- Subject Search -- Keyword

- Any Search -- Keyword

Alternatively, the implementation may conform to the *Z Texas Profile* Functional Area A for Basic Bibliographic Search & Retrieval, with primary focus on library catalogs, level 0, which focuses on recall-oriented search and retrieval.

Enhanced

Participants will make metadata available to be searched using the Z39.50 protocol or its successors or derivatives, including SRU/SRW. The Z39.50 implementation will conform to the *Z Texas Profile* Functional Area C, level 1. This area provides a definition of a core set of searches (and the associated attributes and attribute combinations) required for basic search and retrieval mechanisms needed by library users when interacting with disparate electronic resources, duplicating the Bath Profile Functional Area C. The following elements are used in this area/level:

- Creator Search -- Keyword with Right Truncation
- Title Search -- Keyword with Right Truncation
- Title Search -- Unanchored Phrase
- Subject Search -- Keyword with Right Truncation
- Subject Search -- Unanchored Phrase
- Any Search -- Keyword with Right Truncation
- Any Search -- Unanchored Phrase
- Standard Identifier Search
- Date of Publication Search

Alternatively, the implementation may conform to the *Z Texas Profile* Functional Area A for Basic Bibliographic Search & Retrieval, with primary focus on library catalogs, level 1, which focuses on precision-oriented search and retrieval, or level 2, which defines a variety of more specialized searches, including, for example, searches using controlled vocabularies and specialized author and title searches.

References

- Bath Group. 2003 March. *The Bath Profile: An International Z39.50 Specification for Library Applications and Resource Discovery*. Release 2.0. Retrieved 2 February 2006 from <http://www.collectionscanada.ca/bath/tp-bath2-e.htm>
- NISO. 2003. *Z39.89-2003: U.S. National Z39.50 Profile for Library Applications*. Retrieved 2 February 2006 from http://www.niso.org/standards/standard_detail.cfm?std_id=734
- DLF/NSDL. 2005. *Best Practices for OAI Data Provider Implementations and Shareable Metadata*. Retrieved 2 February 2006 from <http://oai-best.comm.nsd.org/cgi-bin/wiki.pl?TableOfContents>
- Open Archives Forum. 2002 February. *OAI for Beginners: The Open Archives Forum Online Tutorial*. Retrieved 2 February 2006 from <http://www.oaforum.org/tutorial/>
- Open Archives Initiative. 2005 May. *Implementation Guidelines for the Open Archives Initiative Protocol for Metadata Harvesting*. Version 2. Retrieved 2 February 2006 from <http://www.openarchives.org/OAI/2.0/guidelines.htm>
- Open Archives Initiative. 2002 February. *The Open Archives Initiative Protocol for Metadata Harvesting*. Version 2. Retrieved 2 February 2006 from <http://www.openarchives.org/OAI/openarchivesprotocol.html>
- Texas Z39.50 Implementers Group (TZIG). 2004 February. *Z Texas Profile: A Z39.50 Specification for Library Systems Applications in Texas*. Release 3.1. Based on Bath Profile, Release 2.0 and ANSI/NISO Z39.89, Final Draft. Retrieved 2 February 2006 from <http://www.tsl.state.tx.us/ld/projects/z3950/tzigprofilerelease31.html>

Preservation Management Standards

Introduction

Preservation management of digital materials is not a new field. It had its origins in the 1970's, when government agencies were already discovering the rapid rate of obsolescence of digital data. As computer technology has matured, this rate of obsolescence has decreased, but the amount of both "born digital" and digitized materials, and the institutional importance of those materials, has increased dramatically.

In addition to the desire to promote continued accessibility, defined by the Digital Preservation Coalition as the "continued, ongoing usability of a digital resource, retaining all qualities of authenticity, accuracy and functionality deemed to be essential for the purposes the digital material was created and/or acquired for," institutions may be legally required to provide access to both "born digital" and digitized materials under the Texas Public Information Act and the U.S. Freedom of Information Act. This legal mandate for continued access is in sharp contrast to the efforts of the private sector to promote Digital Rights Management (DRM) systems that limit access to digital materials.

Preservation management of digitized materials has largely been derived from preservation management practices for physical materials. This is the origin of the widely-held "scan once" preservation practice. While this may be appropriate for some materials and collections, it is becoming increasingly apparent that digital preservation practices for digitized materials may in some cases require re-digitization of individual items and collections in order to maintain accessibility. It is to be hoped that as the available resolution and color depth of scanning equipment approaches the theoretical limits of photographic processes, the need for redigitization of original source materials will decrease.

There are three common strategies for digital preservation: preservation, emulation, and migration. Preservation typically involves maintaining the original hardware and software used to create the digitized material; for material that is presented via the World Wide Web, this may require maintenance of older browser versions and plug-ins or helper applications. Emulation is a strategy that requires the creation of a "virtual machine" or software tool that will allow older software to run on newer hardware and operating system. It may also include the development of software that maintains backward-compatibility with older formats. Migration is the best and most expensive option; as new formats, software, and hardware appear, materials should be recreated or resaved. Projects working to develop best practices using these various strategies include the Cedars and CAMiLEON projects, both completed, and the MIDESS project.

There are many good sources for information about preservation management. The Northeast Document Conservation Center's *Handbook for digital projects: A management tool for preservation and access* contains excellent overviews and practice information, including information about vendors, though some of the information needs to be updated. The Digital Preservation Coalition, National Library of Australia, and the PADI Gateway regularly maintain and update their *Preservation Management of Digital Material Handbook*, which is also available in the 2001 print edition, written by Maggie Jones and Neil Beagrie.

Levels of Preservation Management

THDI recognizes three levels of standards for preservation management of digitized materials:

Minimal

Participants will use non-proprietary formats whenever possible and will minimize the use of tools for digital rights management (DRM) that may in the future affect the accessibility of their digital collections. Participants will regularly evaluate the accessibility of their collections and will remove, refresh, or recreate collections as they find necessary and practical. Participants will store materials on at least two physical drives, a server and a backup.

Basic

Participants will identify items with medium to long term access requirements and develop preservation management plans for those items and collections. Participants may choose to use the ERPANet "Digital Preservation Policy Tool" in developing those plans. Participants should include plans for adding those materials as appropriate to a trusted digital repository, as defined by the RLG and OCLC 2002 report on attributes of trusted digital repositories, which has been further developed in the 2004 PREMIS Working Group report on implementing preservation repositories. Participants will store materials in at least two media maintained in separate physical locations and will regularly refresh the media as needed to preserve the content, using information and best practices for sustainability of digital formats and media developed by the Library of Congress and the National Digital Information Infrastructure and Preservation Program (NDIIPP).

Enhanced

Participants will meet the requirements of trusted digital repositories, as defined by the RLG and OCLC 2002 report on attributes of trusted digital repositories, which has been further developed in the 2004 PREMIS Working Group report on implementing preservation repositories. Participants will use a digital asset management system that conforms to the OAIS reference model and will develop and use preservation metadata as required by the OAIS reference model as implemented in the PREMIS "Data Dictionary for Preservation Metadata." Participants may choose to become involved in the National Digital Information Infrastructure and Preservation Program (NDIIPP) coordinated by the Library of Congress.

References

- Beagrie, Neil. (2005 Apr 30). Digital Preservation: Best Practice and its Dissemination. *Ariadne* Issue 43. Retrieved March 13, 2006, from <http://www.ariadne.ac.uk/issue43/beagrie/intro.html>
- CAMiLEON project. (2003). Creative Archiving at Michigan & Leeds: Emulating the Old on the New. Retrieved March 15, 2006, from <http://www.si.umich.edu/CAMILEON/>
- Canadian Heritage Information Network (CHIN). (2006). Digital Preservation: Best Practice for Museums. Retrieved March 13, 2006 from http://www.chin.gc.ca/English/Digital_Content/Digital_Preservation/index.html
- Consultative Committee for Space Data Systems. (2002). Reference Model for an Open Archival Information System (OAIS). Adopted by the ISO as ISO 14721:2002. Retrieved March 15, 2006 from <http://ssdoo.gsfc.nasa.gov/nost/wwwclassic/documents/pdf/CCSDS-650.0-B-1.pdf>
- Cedars Project. (2002). Metadata for digital preservation: the Cedars Project outline specification. Retrieved March 15, 2006 from <http://www.leeds.ac.uk/cedars/colman/metadata/metadataspec.html>
- Conway, Paul. (1996). Preservation in the Digital World. Retrieved March 13, 2006 from the CLIR site: <http://www.clir.org/pubs/reports/conway2/index.html>
- Digital Preservation Coalition, National Library of Australia, and the PADI Gateway. (2006). Preservation Management of Digital Material Handbook. Retrieved March 13, 2006 from <http://www.dpconline.org/graphics/handbook/>
- ERPANet. (2003). Digital preservation policy tool. Retrieved March 13, 2006 from <http://www.erpanet.org/guidance/docs/ERPANETPolicyTool.pdf>
- Library of Congress. (2006). National Digital Information Infrastructure and Preservation Program (NDIIPP) website. Retrieved March 15, 2006 from <http://www.digitalpreservation.gov/index.html>
- MIDESS Project. (2005). Management of Images in a Distributed Environment with Shared Services. Retrieved March 15, 2006 from <http://www.leeds.ac.uk/library/midess/>
- National Initiative for a Networked Cultural Heritage (NINCH), and the Humanities Advanced Technology and Information Institute (HATII), University of Glasgow. (2002). The NINCH Guide to Good Practice

- in the Digital Representation and Management of Cultural Heritage Materials: Preservation. Retrieved March 13, 2006 from <http://www.nyu.edu/its/humanities/ninchguide/XIV/>
- National Library of Australia. (2006). Preserving Access to Digital Information (PADI). Retrieved March 13, 2006 from <http://www.nla.gov.au/padi/>
- NDIIPP. Digital formats website. Retrieved March 15, 2006 from <http://www.digitalpreservation.gov/formats/index.shtml>
- OCLC/RLG Working Group on Preservation Metadata. (2002). A metadata framework to support the preservation of digital objects. Dublin, OH: OCLC Online Computer Library, Inc. Retrieved March 15, 2006 from http://www.rlg.org/en/pdfs/pm_framework.pdf
- PREMIS Working Group. (2004 September). Implementing preservation repositories for digital materials: Current practice and emerging trends in the cultural heritage community. Retrieved March 15, 2006 from <http://www.oclc.org/research/projects/pmwg/surveyreport.pdf>
- PREMIS Working Group. (2005). Data dictionary for preservation metadata. Retrieved March 15, 2006 from <http://www.oclc.org/research/projects/pmwg/premis-final.pdf>
- RLG. (2002). OAIS resources. Retrieved March 15, 2006 from http://www.rlg.org/en/page.php?Page_ID=3681
- RLG/OCLC. (2002). Trusted digital repositories: Attributes and responsibilities. Mountain View, CA: RLG. Retrieved March 15, 2006 from <http://www.rlg.org/en/pdfs/repositories.pdf>
- Sitts, Maxine K., ed. (2000). Handbook for digital projects: A management tool for preservation and access. First Edition. Andover, MA: Northeast Document Conservation Center. Retrieved March 13, 2006 from <http://www.nedcc.org/digital/dighome.htm>
- UKOLN. (2006). Good Practice Guide for Developers of Cultural Heritage Web Services: Digital Preservation. Retrieved March 13, 2006 from <http://www.ukoln.ac.uk/interop-focus/gpg/Preservation/>