Texas Heritage Online

Standards & Guidelines for Cultural Heritage Digitization Projects

THDI Steering Committee, 2007.

Revised by THO Standards Working Group, 2011.

Table of Contents

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. Texas Heritage Online (THO), formerly known as 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 and approved by the THDI Steering Committee at the 2007 Annual meeting.

Over the course of the spring and summer of 2010, the THO Standards Work Group (THO-SWG) reviewed the 2007 document. When the original standards were published, THO was not yet online. With the availability of the THO website (www.texasheritageonline.org), the Standards were updated to reflect the framework and minimum requirements for online heritage collections to be discoverable and included in the THO online search. Extensive changes were made to the Preservation Management Standards section, reflecting the many recent developments in this area over the last couple of years. Other changes include updating references to current standards, annotating references, and adding information about standards for audio and video. Lastly, a "Guidelines and Best Practices" section was added for each area to offer additional guidance and resources to readers. These changes were reviewed again by the THO Standards WG Chair and THO Coordinator in the summer of 2011 before presenting the changes to the THO Board.

THO's *Standards & Guidelines 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 or for inclusion in the THO search interface; 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.

Standards & Guidelines for Cultural Heritage Digitization Projects is divided into 5 sections:

1. Digitization
2. Metadata
3. Controlled Vocabulary
4. Interoperability
5. Preservation Management

Within each section, the *Standards* offer a brief introduction to the standard(s) covered in that section, the definitions of "minimal," "basic," and "enhanced," references to published standards, and a section on best practices and other resources.

2011 THO-SWG
Mike Miller, Chair (2008-present)
Austin History Center (AHC)

Amanda York-Focke
Rice University

Daryl Stuhr
Baylor University

Megan Peck
Texas Archive of the Moving Image (TAMI)

THO Coordinator
Danielle Cunniff Plumer
Texas State Library and Archives Commission

THO Board Chair
Michele Reilly
University of Houston

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.

It should be noted that there is no one standard file format for digitizing items, whether they be text documents, photographs, maps, or audio/video files. There may be in some cases a *de facto* standard, such as using the TIFF file format for “master” copies of digitized photographs, that have developed because of common and widespread usage. Likewise, there are no official standards for settings in digitizing items, such as resolution, bit depth, and colors for images, but commonalities do develop through shared best practices. The general guides reference below under resources offer more guidance. More about file formats, etc., will be covered in the Preservation Management Standards.

Levels of Digitization:

THO 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.

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 standard 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.

Standards References

National Information Standards Organization. 2006. *Data Dictionary: Technical Metadata for Digital Still Images* (NISO Z39.87-2006). Approved December 18, 2006 by the American National Standards Institute. http://www.niso.org/kst/reports/standards/kfile_download?id%3Austring%3Aiso-8859-1=Z39-87-2006.pdf&pt=RkGKiXzW643YeUaYUqZ1BFwDhIG4-24RJbcZBWg8uE4vWdpZsJDs4RjLz0t90_d5_ymGsj_IKVa86hjP37r_hKQ00ioOP35W6Q1wkBOLKanPbfamndQa6zkS6rLL3oIr

> From the abstract of this standard: “This standard defines a set of metadata elements for raster digital images to enable users to develop, exchange, and interpret digital image files. The dictionary has been designed to facilitate interoperability between systems, services, and software as well as to support the long-term management of and continuing access to digital image collections.”

Guidelines & Best Practices

BCR’s DCP Digital Imaging Best Practices 2.0 http://www.lyrasis.org/Products-and-Services/Digital-and-Preservation-Services/Digital-Toolbox/Best-Practices-and-Publications.aspx

This was published in June 2008, updating the 1.0 version published in 2003. The Bibliographic Center for Research (BCR) closed in December 2010, with many of its activities and members absorbed by Lyrasis.

Besser, Howard, edited by Sally Hubbard with Deborah Lenert. 2003. *Introduction to Imaging.* Revised Edition. http://www.getty.edu/research/conducting_research/standards/introimages/

From the Getty Institute website: "This primer introduces the technology of digital imaging and outlines many of the challenges faced when creating digital image collections." The book is also available in hard copy.

Cornell University Library. 2003. *Moving Theory Into Practice: Digital Imaging Tutorial*. http://www.library.cornell.edu/preservation/tutorial/.

Federal Agencies Digitization Guidelines Initiative. 2010. http://www.digitizationguidelines.gov/

This is a collaborative effort among federal agencies to define common guidelines and practices for digitizing historical materials. Important documents include their *Digital Imaging Standards* (http://www.digitizationguidelines.gov/stillimages/digstandards.html) and *Technical Guidelines for Digitizing Cultural Heritage Materials* (http://www.digitizationguidelines.gov/stillimages/documents/Technical.html).

University of California at Berkeley. 2007. "imaging best practices." http://www.lib.berkeley.edu/digicoll/bestpractices/image_bp.html

A guide to imaging practices put out by the Digital Publishing Group at the UC Berkeley Libraries. The guide contains information and tutorials on scanning, capture specifications, image standards, file formats, technical metadata, and other topics.

NISO Framework Working Group. 2007. *A Framework of Guidance for Building Good Digital Collections.* 3rd Edition. Baltimore, MD. NISO. http://www.niso.org/publications/rp/framework3.pdf

The 3rd edition (2007) of the *Framework*, published by NISO with support by the Insittute for Museum and Library Services (IMLS), is an online resource for open commenting and feedback. The *Framework* "provides an overview of some of the major components and activities involved in the creation of good digital collections and provides a framework for identifying, organizing, and applying existing knowledge and resources to support the development of sound local practices for creating and managing good digital collections. It is intended for two audiences: cultural heritage organizations planning projects to create digital collections, and funding organizations that want to encourage the development of good digital collections."

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.
http://www.archives.gov/preservation/technical/guidelines.pdf

This document describes guidelines established by NARA in regards to the technical aspects of digitization activities, including image capture, metadata, file formats and naming, and quality control. It is not a proscriptive document in that they do not recommend on single approach over another. This document is being replaced with a new set of guidelines available at: http://www.archives.gov/preservation/products/. This page has a guide called “Reformatting Approaches Based on Original Record Type” that includes information on reformatting a variety of materials, from paper based to video.

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: 2nd Edition*, 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. *Resource Description and Access* (RDA) is a new standard for resource description for digital materials and is based in part on AACR2. 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 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

THO 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.

For inclusion in the THO Search interface, the following metadata fields ***must*** be included:

- Title
- Description/Note
- Date
- Location
- File Type

In addition, the “Creator/Author” field will be displayed in THO search results, but is not required for inclusion. To read definitions for these fields as they are used by THO, refer to the “Input Guidelines for Descriptive Metadata” from the University of North Texas Libraries. This document is available at: http://www.library.unt.edu/digitalprojects/metadata/descriptive-metadata.

Participants will be at the minimal level if they employ any metadata standard that can be crosswalked to include at least the fields mentioned above. THO will assist participating collections in formatting and/or presenting their metadata in such a way that the THO search engine can search and display the necessary information. For more on this, see the THO Interoperability Standards.

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, METS, MODS, TEI or EAD headers, and the Content Standard for Digital Geospatial Metadata (CSDGM), but this list is not meant to be exclusionary. In addition, participants at the basic level will specify a content standard used with their metadata, such as DACS, AACR2, or RDA.

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.

Standards References

Anglo American Cataloging Rules. 2nd Edition, 2005. http://www.aacr2.org/index.html

The Anglo Americna Cataloging Rules (AACR2) is available for purchase in print from ALA. This website provides an overview of the development and governance as well as information on how to purchase or access online.

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.

CCO is a data content standards initiative for the cultural heritage community sponsored by the Visual Resources Association Foundation. The primary focus is on art and architecture, but includes archaeological sites, artifacts, and functional objects. Their website is: http://www.vrafoundation.org/ccoweb/

Dublin Core Metadata Initiative. 2007. *ANSI/NISO Z39.85 - The Dublin Core Metadata Element Set*. http://www.niso.org/kst/reports/standards?step=2&gid=&project_key=9b7bffcd2daeca6198b4ee5a848f9beec2f600e5

This is the NISO Standard, published on May 22, 2007, that defines the 15 metadata elements for resource description.

Dublin Core Metadata Initiative. http://www.dublincore.org/

This is the website devoted to the ongoing discussion and development of the Dublin Core metadata standard. Of particular interest to those working on digitization projects are the “Metadata Basics” and “DCMI Specifications” sections.

Federal Geographic Data Committee. FGDC-STD-001-1998. *Content Standard for Digital Geospatial Metadata (Revised June 1998).* Federal Geographic Data Committee. Washington D.C. http://www.fgdc.gov/metadata/csdgm/

This standard was developed to govern the development, use, sharing and dissemination of geographic data.

Joint Steering Committee for Development of RDA (JSC). 2010. *RDA: Resource Description and Access.* http://www.rdatoolkit.org/

The RDA standard is available via subscription from the above URL. A full text print version of the RDA standard is available for purchase directly from ALA (see: http://www.alastore.ala.org/detail.aspx?ID=3065). Also, the JSC publishes a website with more information about this standard at: http://www.rda-jsc.org/index.html.

Library of Congress. *Encoded Archival Description (EAD): Official EAD 2002 Web Site*. http://www.loc.gov/ead/

EAD is a standard for encoding archival finding aids using Extensible Markup Language (XML). The standard is maintained by the LOC in partnership with the Society of American Archivists. The current version (2002) is closely tied to the content standard, Describing Archives: A Content Standard (DACS).

Library of Congress. *MADS: Metadata Authority Descriptive Schema.* http://www.loc.gov/standards/mads/

MADS is an XML schema for authority control elements and is a companion to MODS. The current version as of June 2010 is 1.0 (http://www.loc.gov/standards/mads/mads.xsd)

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.

The full version of this document, including updates through Update 11 (February 2010), is available at http://www.loc.gov/marc/bibliographic/

Library of Congress. *METS: Metadata Encoding & Transmission Standard Official Web Site*. http://www.loc.gov/standards/mets/

METS is a schema for encoding descriptive, administrative, and structural metadata about objects in a digital library using XML. The METS schema is in version 1.9 as of February 2010 (http://www.loc.gov/standards/mets/mets-schemadocs.html).

Library of Congress. *MODS: Metadata Object Description Schema.* http://www.loc.gov/standards/mods/

MODS is a bibliographic metadata schema used for a variety of purposes and is intended to complement other metadata standards, such as METS and MARC 21. The MODS schema is in version 3.4 as of June 2010 (http://www.loc.gov/standards/mods/v3/mods-3-4.xsd)

Society of American Archivists. 2007. *Describing Archives: A Content Standard.* Chicago: Society of American Archivists

Describing Archives: A Content Standard (DACS) is a set of rules for describing archives, personal papers, and manuscript collections. It replaces Archives, Personal Papers, and Manuscripts (APPM). More information can be found online at: http://www.archivists.org/governance/standards/dacs.asp

Guidelines & Best Practices

Baca, Murtha (ed.), Tony Gill, Anne J. Gilliland, and Mary S. Woodley. 2008. *Introduction to Metadata*. Los Angeles: Getty Research Institute. http://www.getty.edu/research/conducting_research/standards/intrometadata/

"Best Practices for CONTENTdm users creating shareable metadata." Draft 1.9. Available at: http://contentdmmwg.wikispaces.com/file/view/BPG1.9.pdf

CDP Metadata Working Group. 2006 September. *Dublin Core Metadata Best Practices*, Version 2.1.1. Available at from http://www.lyrasis.org/Products-and-Services/Digital-and-Preservation-Services/Digital-Toolbox/Best-Practices-and-Publications.aspx.

NISO. 2004. *Understanding Metadat*a. Bethesda, MD: NISO Press. http://www.niso.org/standards/resources/UnderstandingMetadata.pdf

OCLC/RLG Working Group on Preservation Metadata. 2002. *Preservation Metadata and the OAIS Information Model: A Metadata Framework to Support the Preservation Objects*. Dublin, OH. Online Computer Library Center, Inc. http://www.oclc.org/research/activities/past/orprojects/pmwg/pm_framework.pdf

This report is the result of a joint OCLC/RLG study to create a consensus in the area of preservation metadata. See the Preservation Management Standards section for more information.

Controlled Vocabulary

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 **taxonomies** or **structured classifications**, which generally create hierarchical and/or faceted relationships among terms. They may be expressed as formal **ontologies**, modeling objects and/or concepts and the relationships between them.

In traditional library and museum practice, controlled vocabularies have been used primarily as classification tools (DDC, LCC, Chenhall's Nomenclature) and for designation of subject headings (LCSH). New work in this area ties controlled vocabularies into development of the Semantic Web, defined as "a common framework that allows data to be shared and reused across application, enterprise, and community boundaries" (W3C Semantic Web Activity).

Properly developed controlled vocabulary supports both the search and browse functions of information retrieval. Many modern content management systems such as Drupal and SharePoint can be configured with taxonomies that constrain the overall architecture of the system to improve retrieval. While uncontrolled keywords ("tags") remain useful tools for description (and are particularly useful in the area of user-contributed metadata, or folksonomies), they should be considered a supplement to controlled vocabularies rather than a replacement.

The Getty Museum's *Introduction to Controlled Vocabularies: Terminology for Art, Architecture, and Other Cultural Works* provides a good overview of controlled vocabulary.

Levels of Controlled Vocabulary

Texas Heritage Online recognizes three levels of controlled vocabulary use:

Minimal

Participants will use values from controlled vocabularies for the following three metadata elements:

> **Language.** Use values from a standard source, such as IETF RFC 5646 (http://tools.ietf.org/html/rfc5646), the MARC Code List for Languages (http://www.loc.gov/marc/languages/) or the UNT Vocabularies (http://digital2.library.unt.edu/vocabularies/languages/).
>
> **Resource Type.** Use values from a standard source, such as the DCMI Type Vocabulary (http://dublincore.org/documents/dcmi-type-vocabulary/), RDA Vocabularies Carrier Type (http://metadataregistry.org/concept/list/vocabulary_id/46.html), or the UNT Vocabularies (http://digital2.library.unt.edu/vocabularies/resource-types/).

Coverage (spatial). Use values from a standard source, such as the Getty Thesaurus of Graphic Names (http://www.getty.edu/research/tools/vocabulary/tgn/index.html). For projects limited to materials created within the state of Texas, include the modern form of the county name followed by the word "County" (e.g., Travis County), using a tool such as the *Atlas of Historical County Boundaries* (http://publications.newberry.org/ahcbp/statepages/Texas.html) when needed to determine the modern county.

Participants may also use locally developed and maintained controlled vocabularies 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

In addition to meeting minimal requirements, 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:

- Library of Congress Subject Headings (LCSH) http://id.loc.gov/authorities/
- Medical Subject Headings (MeSH) http://www.nlm.nih.gov/mesh/2002
- Thesaurus for Graphic Materials http://id.loc.gov/vocabulary/graphicMaterials.html
- Art and Architecture Thesaurus (AAT) http://www.getty.edu/research/tools/vocabularies/aat/
- Getty Thesaurus of Geographic Names (TGN) http://www.getty.edu/research/tools/vocabulary/tgn/index.html
- Geographic Names Information System (GNIS) http://geonames.usgs.gov/
- Library of Congress Authorities http://authorities.loc.gov/
- Union List of Artist Names (ULAN) http://www.getty.edu/research/tools/vocabularies/ulan/
- RDA Vocabularies http://metadataregistry.org/rdabrowse.htm

Enhanced

In addition to meeting minimal and basic requirements, participants will participate in the development of controlled vocabulary appropriate to their projects, regions, and resources either by contributing to established standards, joining collaborative projects such as the Getty Cultural Objects Name Authority project (http://www.getty.edu/research/tools/vocabularies/cona/index.html), or by sharing their local vocabularies online using an established schema such as EAC-CPF or a formal ontology expressed in RDF, SKOS, OWL, etc.

Standards References

Library of Congress. 2011. *Extended Date-Time Format* (draft specification). http://www.loc.gov/standards/datetime/spec.html

National Information Standards Organization (U.S.). 2005. *Guidelines for the construction, format, and management of monolingual controlled vocabularies*. (ANSI/NISO Z39.19-2005). http://www.niso.org/kst/reports/standards?step=2&gid%3Austring%3Aiso-8859-1=&project_key%3Austring%3Aiso-8859-1=7cc9b583cb5a62e8c15d3099e0bb46bbae9cf38a

Society of American Archivists and Berlin State Library. 2010. *Encoded Archival Context: Corporate Bodies, Persons, and Families* (EAC-CPF). http://eac.staatsbibliothek-berlin.de/

Guidelines & Best Practices

The preferred Texas Heritage Online authority list for local and regional terms not included in one of the standard sources of controlled vocabulary listed above is *The Handbook of Texas Online* (http://tshaonline.org/handbook/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. *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.

Patricia Harpring. 2010. *Introduction to Controlled Vocabularies: Terminology for Art, Architecture, and Other Cultural Works*. Los Angeles: Getty Publications. http://www.getty.edu/research/publications/electronic_publications/intro_controlled_vocab/index.html

Library of Congress. "Authorities and Vocabularies." http://id.loc.gov/

> The Library of Congress has opened its ID.LOC.GOV web service, Authorities and Vocabularies, with the Library of Congress Subject Headings (LCSH) as the initial offering. The primary goal of this service is to enable machines to programmatically access data at the Library of Congress but the web interface also provides simple user access. The site now offers, in addition to LCSH, MARC list for Relators, Geographic Terms, Languages, and Countries, ISO 639 standards for languages, Cryptographic Hash Functions, Thesaurus of Graphic Terms, Preservation Events, and Preservation Level Role.

W3C. 2009. *OWL 2 Web Ontology Language Document Overview*. http://www.w3.org/TR/owl2-overview/

W3C. 2004. *Resource Description Framework (RDF): Concepts and Abstract Syntax*. http://www.w3.org/TR/2004/REC-rdf-concepts-20040210/

W3C. 2009. *SKOS Simple Knowledge Organization System Reference*. http://www.w3.org/TR/skos-reference/

Interoperability

Introduction

Interoperability refers to the capacity of a system to interact with other systems in a predefined manner. Texas Heritage Online does not intend to create a single state-wide repository for all of Texas; rather, Texas Heritage Online 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 Texas Heritage Online to specify particular interoperability requirements and standards.

Currently, there are three ways to provide a common search interface for online collections of cultural heritage materials. 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 and so should be supported by all institutions. The second approach involves the use of a common query language and an application programming interface (API) to search disparate resources in real-time. In libraries, the most common API in use is the Z39.50 protocol for searching across library catalogs; the OpenSearch and SRU protocols (used by WorldCat, among others) are standard APIs for online resources, but many systems use proprietary or custom APIs, making them difficult, if not impossible, for Texas Heritage Online to support. The third approach involves the use of a metadata harvester that accesses pre-identified collections known to have been developed according to the Open Archives Initiative's Protocol for Metadata Harvesting. Most dedicated digital asset management systems, including CONTENTdm, DSpace, Fedora, and Omeka, support this protocol, and some content management systems, including Drupal, can be configured to support it.

Interoperability is not limited to systems, however, and systems or syntactic interoperability is perhaps the lowest level of interoperability; true semantic interoperability is a much loftier goal. The World Wide Web's Semantic Web initiative is building interest in other ways to share structured metadata, particularly in the form of linked data, and this is an area of emerging interest in the cultural heritage community.

Levels of Interoperability

Texas Heritage Online recognizes three levels of interoperability:

Minimal

Participants will make metadata available online in a manner that supports indexing by Web search engines. Where metadata is stored in databases or other systems that do not permit indexing directly, participants will make a sitemap available following the protocol established at http://www.sitemaps.org/. Note that this metadata cannot currently be indexed by Texas Heritage

Online, but selected collections meeting this requirement may be listed as "Additional Resources" at TexasHeritageOnline.org, depending on their scope and significance.

Basic

In addition to meeting the minimal requirements, participants will make their collections available using an OAI-PMH compliant system. The system should provide complete metadata records, including thumbnail information, in the oai_dc metadata format, whether or not such information is displayed in the normal interface.

Enhanced

In addition to meeting both the minimal and basic requirements, participants will expose metadata using other standard and/or experimental APIs, such as SRU and OpenSearch, and will share information about the available APIs to promote date reuse. Participants may also choose to make structured metadata available directly, expressed in schema such as the microdata schema listed at http://www.schema.org/ or as RDF or RDFa.

Standards References

Library of Congress (2012). Search/Retrieve via URL (SRU), Version 1.2. http://www.loc.gov/standards/sru/specs/index.html

Open Archives Initiative. 2002 February. *The Open Archives Initiative Protocol for Metadata Harvesting*, Version 2. http://www.openarchives.org/OAI/openarchivesprotocol.html

Guidelines & Best Practices

There are many best practices documents that describe configuring OAI-PMH support for specific systems. Institutions that do not run an OAI-compliant system may wish to look at the list of tools available at http://www.openarchives.org/pmh/tools/tools.php; newer tools not included on that page include the OAI2 module for Drupal (http://drupal.org/project/oai2), the eXtensible Catablog project's OAI Toolkit (http://code.google.com/p/xcoaitoolkit/), and the University of Michigan's OAI Toolkit (http://sourceforge.net/projects/umoaitoolkit/). Additional best practices documents include, but are not limited to, the following list.

A9.com, Inc. (2005). *OpenSearch.org*, Version 1.1. http://www.opensearch.org/Home

DCMI. 2009. *Interoperability Levels for Dublin Core Metadata*. http://www.dublincore.org/documents/2009/05/01/interoperability-levels/

> From the Introduction: The evolving assumptions which over the past decade have led from fifteen elements [RFC2413] to the Singapore Framework for Dublin Core Application Profiles [DCAP] can be captured in a layered model of interoperability. The model of levels presented here addresses the need felt in many communities to position

various projects with various degrees of interoperability with Dublin Core but lacking an appropriate terminology

DLF. 2006. *Best Practices for OAI Data Provider Implementations*. http://webservices.itcs.umich.edu/mediawiki/oaibp/index.php/DataProviderPractices

Google, Inc., Yahoo, Inc., and Microsoft Corporation (2011). *Schema.org*. http://www.schema.org/docs/terms.html

Google, Inc., Yahoo, Inc., and Microsoft Corporation (2008). *Sitemaps.org*. http://www.sitemaps.org/

NISO Framework Advisory Group. 2007. *A Framework of Guidance for Building Good Digital Collections*. 3rd edition. Bethesda, MD: National Information Standards Organization. http://framework.niso.org/

OCLC Digital Collection Services. 2010. *Best Practices for CONTENTdm and other OAI-PMH compliant repositories*. http://www.oclc.org/gateway/support/best_practices.pdf

Elings, Mary W. and Günter Waibel. 2007. "Metadata for All: Descriptive Standards and Metadata Sharing across Libraries, Archives and Museums." *First Monday* (12:3): http://firstmonday.org/issues/issue12_3/elings/index.html

Heath, Tom and Christian Bizer. 2011. *Linked Data: Evolving the Web into a Global Data Space* (1st edition). Synthesis Lectures on the Semantic Web: Theory and Technology, 1:1, 1-136. Morgan & Claypool. http://linkeddatabook.com/editions/1.0/

Open Archives Forum. 2002. *OAI for Beginners: The Open Archives Forum Online Tutorial*. http://www.oaforum.org/tutorial/

Shreeves, Sarah L., Jenn Riley, and Liz Milewicz. August 2006. "Moving towards shareable metadata." *First Monday* (11:8). http://firstmonday.org/issues/issue11_8/shreeves/index.html

TIPR: Towards Interoperable Preservation Repositories. 2011. *Repository eXchange Package (RXP) Spec.* version 0.96. http://wiki.fcla.edu:8000/TIPR/21

From their website: "The Repository Exchange Package (RXP) is a hierarchical packaging format designed to facilitate the exchange of Archival Information Packages (AIPs) between digital repositories." The TIPR project, funded by an IMLS grant, ran October 2008 to September 2011 and developed a model for repository-to-repository transfer.

W3C Library Linked Data Incubator Group. 2011. *LLD XG Final Report* (draft). http://www.w3.org/2005/Incubator/lld/wiki/DraftReportWithTransclusion

Preservation Management

Introduction

Preservation management of digital materials is not a new field. It had its origins in the 1970s 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, have 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.

A crucial partner to any of these three strategies is the concept of "lots of copies keeps stuff safe" and that those copies need to stored in a geographically diverse manner. Recent attention to distributed digital preservation networks is addressing this need and providing information on ways to implement these ideas at various levels.

See Guidelines & Best Practices below for more detailed information.

Levels of Preservation Management

THO 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, preferably a server which is backed up to tape and spinning hard drive, located in different places.

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 PLATO (PLANETS Preservation Planning Tool) http://www.ifs.tuwien.ac.at/dp/plato/intro.html in developing those plans. Participants should include plans for adding those materials as appropriate to a trusted digital repository, as defined by the **Trustworthy Repositories Audit and Certification (TRAC): Criteria and Checklist** http://www.crl.edu/sites/default/files/attachments/pages/trac_0.pdf, published by the Center for Research Libraries (CRL). 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 **Trustworthy Repositories Audit and Certification (TRAC): Criteria and Checklist** http://www.crl.edu/sites/default/files/attachments/pages/trac_0.pdf, published by the CRL. 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.

Standards References

Consultative Committee for Space Data Systems. 2002. *Reference Model for an Open Archival Information System (OAIS): Recommendation for Space Data System Standards.* Washington, D.C. National Aeronautics and Space Administration. http://public.ccsds.org/publications/archive/650x0b1.pdf

Prepared by the Consultative Committee for Space Data Systems (CCSDS), this document is "a technical Recommendation for use in developing a broader consensus on what is required for an archive to provide permanent, or indefinite long-term, preservation of digital information. This Recommendation establishes a common framework of terms and concepts which comprise an Open Archival Information System (OAIS)." [description from: *http://www.niso.org/news/events/2010/preservation/resources]*

International Standards Organization. *Technical Report: Information Technology – Multimedia Framework (MPEG-21).* ISO/IEC 21000-1:2004(E).

MPEG-21 aims at defining a normative albeit open framework for multimedia creation and sharing for use by all the players in the delivery and consumption chain. This open framework will provide content creators and service providers with equal opportunities in the MPEG-21 enabled open market. This will also be to the benefit of the content consumer providing them access to a large variety of content in an interoperable manner. The MPEG-21 vision can thus be summarized as follows: *to define a multimedia framework to enable transparent and augmented use of multimedia resources across a wide range of networks and devices used by different communities*. A copy of this report, and the accompanying parts to the standard, is available for a free download at: http://standards.iso.org/ittf/PubliclyAvailableStandards/index.html

ISO 28500:2009. Information and documentation – WARC file format. http://www.iso.org/iso/catalogue_detail?csnumber=44717

WARC (Web ARChive) is a file format for Internet archiving. It “offers a convention for concatenating multiple data objects into one long file. The format can be used to build applications for harvesting, managing, accessing and exchanging content.”

Library of Congress. 2010. PREMIS: Preservation Metadata. http://www.loc.gov/standards/premis/

PREMIS stands for “PREservation Metadata: Implementation Strategies.” The PREMIS Data Dictionary defines preservation metadata for various repositories to support the digital preservation process.

Library of Congress. 2010. TextMD: Technical Metadata for Text. http://www.loc.gov/standards/textMD/

“textMD” is a XML Schema that details technical metadata for text-based digital objects. It most commonly serves as an extension schema used within the Metadata Encoding and Transmission Schema (METS) administrative metadata section. However, it could also exist as a standalone document.”

National Information Standards Organization. 2006. *ANSI/NISO Z39.87-2006: Data Dictionary – Technical Metadata for Digital Still Images.* Bethesda, MD. National Information

Standards Organization. http://www.niso.org/kst/reports/standards?step=2&gid=None&project_key%3Austring%3Aiso-8859-1=b897b0cf3e2ee526252d9f830207b3cc9f3b6c2c

"This standard defines a set of metadata elements for raster digital images to enable users to develop, exchange, and interpret digital image files. The dictionary has been designed to facilitate interoperability between systems, services, and software as well as to support the long-term management of and continuing access to digital image collections." [from the Abstract].

ODF Alliance. Open Document Format. http://www.odfalliance.org/

OpenDocument Format (ODF) is the only open standard for office applications, and it is completely vendor neutral. Developed through a transparent, multi-vendor/multi-stakeholder process at OASIS (Organization for the Advancement of Structured Information Standards - http://www.oasis-open.org/home/index.php), ODF is an open, XML-based document file format for displaying, storing and editing office documents, such as spreadsheets, charts, and presentations. It is available for implementation and use free from any licensing, royalty payments, or other restrictions. In May 2006, it was approved unanimously as an International Organization for Standardization (ISO) and International Electrotechnical Commission (IEC) standard (ISO/IEC 26300).

PDF/A Competence Center. 2010. http://www.pdfa.org/doku.php

PDF/A is an ISO standard (ISO-19005-1 - Document management - Electronic document file format for long-term preservation - Part 1: Use of PDF 1.4 (PDF/A-1)) for using PDF format for the long-term archiving of electronic documents.

Guidelines & Best Practices

In addition to using the information provided below, organizations may want to consider joining a national consortium for the preservation of digital objects. Examples of these consortia include LOCKSS, the MetaArchive Cooperative, and Chronopolis.

Beagrie, Neil. April 30, 2005. "Digital Preservation: Best Practice and its Dissemination." *Ariadne* Issue 43. http://www.ariadne.ac.uk/issue43/beagrie/intro.html

This article discusses the development of the Digital Preservation Handbook, originally authored by Beagrie and now maintainied by the Digital Preservation Coalition (see their entry below for more on the handbook).

Brown, Adrian. 2003. *Digital Preservation Guidance Note: Selecting File Formats for Long-Term Preservation.* DPGN-01. London, UK. The National Archives. http://www.nationalarchives.gov.uk/documents/selecting_file_formats.pdf

This is part of a series of documents giving advice to records creators on the preservation and management of electronic records in the British government. This title gives general advice on selecting file formats for storage of electronic records.

Canadian Heritage Information Network (CHIN). 2009. Digital Preservation: Best Practice for Museums. http://www.pro.rcip-chin.gc.ca/sommaire-summary/preservation_numerique-digital_preservation-eng.jsp

This is one of the few resources to look at digitization and digital asset management from a purely museum viewpoint.

Caplan, Priscilla. 2006. “Preservation Metadata.” *Curation Reference Manual*. Edinburgh. Digital Curation Centre. http://www.dcc.ac.uk/resources/curation-reference-manual/completed-chapters/preservation-metadata

"This installment explores some of the issues involved in recording and managing preservation metadata over time, provides a summary of current activity, and examines some of the key issues relating to preservation metadata including provenance, file format specification and packaging."

Center for Research Libraries. CRL Certification and Assessment of Digital Repositories. http://www.crl.edu/archiving-preservation/digital-archives/certification-and-assessment-digital-repositories

This Center for Research Libraries page is the home for information on a project initiated in 2009 to conduct "in-depth assessments of two repositories of interest to the CRL community: Portico and HathiTrust. The purpose of these was to promote understanding of and, where justified, confidence in, digital repositories."

Dance Heritage Coalition. Digital Video Preservation Reformatting Project. http://www.danceheritage.org/preservation/DigitalVideoPreservation1.pdf

Digital Curation Centre. 2010. *DRAMBORA (Digital Repository Audit Method Based On Risk Assessment) Interactive.* http://www.repositoryaudit.eu/

This online toolkit, from the Digital Curation Centre (DCC) and DigitalPreservationEurope (DPE), is "intended to facilitate internal audit by providing repository administrators with a means to assess their capabilities, identify their weaknesses, and recognise their strengths." [description from: *http://www.niso.org/news/events/2010/preservation/resources*].

Digital Preservation Coalition. 2009. *Digital Preservation Handbook.* Heslington, York. Digital Preservation Coalition. http://www.dpconline.org/graphics/handbook/

The handbook provides an internationally authoritative and practical guide to the subject of managing digital resources over time and the issues in sustaining access to them.

Educopia Institute. MetaArchive Cooperative. http://www.metaarchive.org/

> From their website: “The MetaArchive Cooperative provides low-cost, high-impact preservation services to help ensure the long-term accessibility of the digital assets of universities, libraries, museums, and other cultural heritage institutions.”

Electronic Resource Preservation and Access Network (ERPANet). 2003. *Digital Preservation Policy Tool.* http://www.erpanet.org/guidance/docs/ERPANETPolicyTool.pdf

> This tool examines policies in use or in project for preserving and maintaining digital materials and ensuring their availability for current and future use.

Frost, Hannah. 2008. “Audio Preservation.” CoOl: Conservation Online. http://cool.conservation-us.org/bytopic/audio/

> A resource guide for audio preservation, including a section on digital audio.

Frost, Hannah. 2011. “Video Preservation.” CoOl: Conservation Online. http://cool.conservation-us.org/bytopic/video/

> A resource guide for video preservation, including a section on digital audio.

Global Digital Formats Registry. Unified Digital Formats Registry (UDFR). http://www.gdfr.info/udfr.html

> A registry to "support the requirements and use cases of the larger community compiled for GDFR [Global Digital Format Registry] and would be seeded with PRONOM's software and formats database."

HathiTrust: A Shared Digital Repository. http://www.hathitrust.org/

> “HathiTrust was conceived as a collaboration of the thirteen universities of the Committee on Institutional Cooperation and the University of California system to establish a repository for these universities to archive and share their digitized collections."

InterPARES Project: International Research on Permanent Authentic Records in Electronic Systems. 2010. http://www.interpares.org/

> The InterPARES (The International Research on Permanent Authentic Records in Electronic Systems) Project "aims at developing the knowledge essential to the long-term preservation of authentic records created and/or maintained in digital form and providing the basis for standards, policies, strategies and plans of action capable of ensuring the longevity of such material and the ability of its users to trust its authenticity." [description from: *http://www.niso.org/news/events/2010/preservation/resources*].

Lavoie, Brian and Richard Gartner. 2005. *Technology Watch Report: Preservation Metadata.* DPC Technology Watch Series Report 05-01.

Library of Congress. 2010. MIX: NISO Metadata for Images in XML Schema. http://www.loc.gov/standards/mix//

The Library of Congress is developing an XML schema, called MIX, to provide a format for data specified in this NISO Standard data dictionary. MIX is in version 2.0.

Library of Congress. 2010. National Digital Information Infrastructure and Preservation Program (NDIIPP) http://www.digitalpreservation.gov/index.html

The mission of the National Digital Information Infrastructure and Preservation Program is to develop a national strategy to collect, preserve and make available significant digital content, especially information that is created in digital form only, for current and future generations.

Library of Congress. 2010. Sustainability of Digital Formats: Planning for Library of Congress Collections. http://www.digitalpreservation.gov/formats/

This website provides information about digital content formats.

The National Archives. PRONOM Technical Registry. http://www.nationalarchives.gov.uk/PRONOM/Default.aspx

From the UK National Archives, PROMOM is an online registry of technical information that serves as a "resource for anyone requiring impartial and definitive information about the file formats, software products and other technical components required to support long-term access to electronic records and other digital objects of cultural, historical or business value."

National Initiative for a Networked Cultural Heritage (NINCH). 2002. *The NINCH Guide to Good Practice in the Digital Representation and Management of Cultural Heritage Materials.* http://www.nyu.edu/its/humanities/ninchguide/XIV/

From the report: “The National Initiative for a Networked Cultural Heritage (NINCH) is a US-based coalition of some 100 organizations and institutions from across the cultural sector: museums, libraries, archives, scholarly societies, arts groups, IT support units and others. It was founded in 1996 to ensure strong and informed leadership from the cultural community in the evolution of the digital environment. Our task and goal, as a leadership and advocacy organization, is to build a framework within which these different elements can effectively collaborate to build a networked cultural heritage.”

National Library of Australia. 2010. Preserving Access to Digital Information (PADI). http://www.nla.gov.au/padi/

From their website: “The National Library of Australia's Preserving Access to Digital Information (PADI) initiative aims to provide mechanisms that will help to ensure that information in digital form is managed with appropriate consideration for preservation and future access.”

PLANETS: Preservation and Long-term Access through Networked Services. 2010. http://www.planets-project.eu/

PLANETS is a "four-year project co-funded by the European Union under the Sixth Framework Programme to address core digital preservation challenges. The primary goal for Planets is to build practical services and tools to help ensure long-term access to our digital cultural and scientific assets." [description from: *http://www.niso.org/news/events/2010/preservation/resources]*

Planets Team, Universität zu Köln. About XCL - eXtensible Characterization Language. http://planetarium.hki.uni-koeln.de/planets_cms/about-xcl

From the site: "One of the goals of the preservation characterisation sub-project is to help develop a way to monitor whether a file format migration from format A to format B has fulfilled this condition. ...The Planets team of the Universität zu Köln (UzK) is developing ways to describe these file formats in a way to make the comparison of the information contained within files in different formats possible. This is done with two formal languages, called the Extensible Characterisation Definition Language (XCDL) and the Extensible Characterisation Extraction Language (XCEL), which describe formats and the information contained within individual files."

PLATO: PLANETS Preservation Planning Tool. 2010. http://www.ifs.tuwien.ac.at/dp/plato/intro.html

Plato is "a decision support tool that implements a solid preservation planning process and integrates services for content characterisation, preservation action and automatic object comparison in a service-oriented architecture to provide maximum support for preservation planning endeavours." [description from: *http://www.niso.org/news/events/2010/preservation/resources]*. Version 2.1 was released November 2009.

Preservation of Digital Audiovisual Content. Briefing Paper. Digital Preservation Europe http://www.digitalpreservationeurope.eu/publications/briefs/audiovisual_v3.pdf

A brief overview of the status of the preservation of digital AV content.

San Diego Supercomputing Center, University of San Diego. Chronopolis: Preserving Our Digital Heritage. http://chronopolis.sdsc.edu/

Chronopolis is a national center for the management, long-term preservation, and promulgation of national digital assets.

Stanford University. LOCKSS (Lots of Copies Keeps Stuff Safe). http://lockss.stanford.edu/lockss/Home

> From their website: “LOCKSS (Lots of Copies Keep Stuff Safe), based at Stanford University Libraries, is an international community initiative that provides libraries with digital preservation tools and support so that they can easily and inexpensively collect and preserve their own copies of authorized e-content.”

Swiss Federal Archives. SIARD (Software Independent Archiving of Relational Databases) Format. http://www.bar.admin.ch/dienstleistungen/00823/00825/index.html?lang=en

> SIARD is an open, published, standardised archiving format for relational databases. It provides a set of software tools, SIARD Suite, to convert databases into the SIARD format. SIARD format and SIARD Suite offer a unique archival solution to preserve and access database content including metadata2 and relations over the long-term.

Trustworthy Repositories Audit and Certification (TRAC): Criteria and Checklist. Chicago. The Center for Research Libraries. 2007. http://www.crl.edu/sites/default/files/attachments/pages/trac_0.pdf

> Published by the National Archives and Records Administration (NARA). From the document: "Regardless of size or purpose, all repositories should be encouraged to use this checklist as a tool for objective evaluation whether it is accomplished in-house or by an objective, third-party auditor, and regardless of whether it is accomplished for local information gathering and evaluation or as a part of an international or national certification process." [description from: *http://www.niso.org/news/events/2010/preservation/resources*].

Gudielines & Best Practices - Video

The information presented here is merely a set of suggestions and guidelines, as the field of digital video preservation is still rapidly evolving. Even the Library of Congress and NARA have disclaimers on their digital a/v collections stating that due to a lack of standardization, they are using popular and emergent formats.

Preserving digital video files presents the archivist with a conflict between what is traditionally considered best practice and what is practical or even possible. While archival convention informs us that a preservation copy should be uncompressed, or at least compressed using a lossless codec, significant practical issues arise. Uncompressed files can become prohibitively large; not only will they fill massive amounts of storage space, they also require increased computer power to render the video when it is being watched. To give an idea of the numbers in question, one frame of uncompressed video uses approximately 1MB of storage space, or nearly 30MB per second. This adds up to roughly 1GB per 45 seconds. The application of a standard DV codec changes the ratio to about 1GB per five minutes.

Further complicating the question of using uncompressed files, many files, whether captured from analog materials or born digital, are created in a compressed format.

Minimal

Because of the evolving nature of online video formats, at this time having only an online access version is not acceptable; basic is the lowest level recommended.

With online file formats rapidly evolving, predicting which ones will be widely accepted in a few years is a gamble. Should the institution's chosen format fall out of wide usage, the institution will need to make a new online version. Transcoding, or applying multiple codecs to the same file, in this case, from one access version to another, is not recommended. The loss of data and quality are unpredictable, and the results can be wholly unsatisfactory.

Basic

Participants create a "master" file in a widely accepted, non-proprietary, publicly open format, and an online access copy. In an ideal scenario, a lossless codec will be used, but when file size is a concern, a widely accepted, if lossy, codec can be used. Participants should keep their collections uniform by applying the chosen codec and file type to all objects, regardless of the origin of the digital file.

Enhanced

In addition to online access copies, participants will create "master" digital objects, and keep them in their captured "RAW" format. Participants will create "preservation" copies using a lossless format such as MJ2K.