![DIGITAL LIBRARIES University Libraries UNT]

UNT Libraries' Digital Preservation Policy Framework

Date: October 2015

Version: 1.0

Contributors:

Mark Phillips Assistant Dean for Digital Libraries
Ana Krahmer Supervisor, Digital Newspaper Unit
Hannah Tarver Department Head, Digital Projects Unit
Daniel Alemneh Supervisor, Digital Curation Unit
Laura Waugh Repository Librarian for Scholarly Works

Note: This is an archived version of this document.
An up-to-date version if available at: http://www.library.unt.edu/policies/

UNT Libraries' Digital Preservation Policy Framework

1. Scope

This policy addresses preservation of digital collections and resources for which the University of North Texas Libraries (UNTL) is the primary custodian. Although this policy only addresses digital collections and resources for which UNTL is the primary custodian, UNTL, to the best of its abilities, has responsibility for informing, consulting, and as appropriate coordinating with other units of the University of North Texas to ensure that UNT faculty, staff, and students will have adequate ongoing access to administrative, scholarly, and other digital resources created at UNT outside of the UNT Libraries. Further, UNTL personnel will also work externally through consortia (e.g., the Texas Digital Library (TDL), and Cross Timbers Library Collaborative (CTLC)), licensing agreements, etc. to ensure that UNT faculty, staff, and students will have adequate continuing access to all currently available digital resources locally owned and managed by the UNTL. UNTL, however, cannot guarantee preservation for materials that it does not own and manage.

2. Purpose

This document formalizes the UNTL's continued commitment to the long-term stewardship for, preservation of, and sustainable access to its diverse and extensive range of digital assets. In alignment with the UNTL mission to create, acquire, organize, disseminate, and preserve digital content, this policy makes explicit UNTL's long-term commitment to the University of North Texas (UNT) community as its trusted digital repository. The UNTL's digital stewardship efforts contribute to UNT's mission to:
- Build a world-class faculty
- Develop academic programs that define UNT as a leader in scholarship and artistic endeavors
- Improve the quality of the teaching and learning environment
- Enhance and better serve the student body
- Create a more diverse university community
- Help build Texas' future by ensuring access to this corpus of information over time

3. Objectives

The primary purpose of digital stewardship and preservation is to collect and maintain the intellectual and cultural heritage important to UNT, while at the same time making sure that it is accessible and held in trust for future use. The objectives in this statement define a framework to:

- Identify, through systematic selection, digital assets to be preserved across new generations of technology.
- Maintain continuous access to reliable data at bit-stream level, the digital assets encoded in the bit streams, as well as access to the intended contextual and intellectual meaning of the digital assets.
- Include in the scope of the program materials that originated in digital form and those that were converted to digital form.
- Protect UNTL's digital investments through a fully-implemented digital preservation program.
- Demonstrate organizational commitment through the identification of sustainable strategies.
- Develop a cost-effective program through means such as system-wide integration, shared responsibilities, and automating human-intensive efforts, when possible.
- Comply with prevailing community standards for digital preservation and access.
- Seek, expand, and develop digital preservation methods that are appropriate for UNT.

4. Mandate

UNTL's mandate for digital preservation is five-fold:
- **Scholarship :** As an institution of higher education, UNT is obligated to support core functions such as scholarship, teaching, and learning. As more resources and services associated with these functions become digital, UNTL's responsibilities must expand to include the identification, stewardship, and preservation of designated digital content.
- **Institutional records:** UNT has charged UNTL with maintaining the University Archives by collecting and preserving university records that best document the history of UNT, including those in electronic format.
- **Legal obligations:** UNT has mandated responsibilities to preserve and maintain access to certain digital collections, as well as responsibilities as a federal depository library. Some legal obligations derived from Federal and State laws require us to maintain files in an archival fashion.
- **Organizational commitment:** UNTL's commitment to digital preservation is explicitly cited in the UNTL's current strategic plan, which calls for UNTL to:
  - develop and implement a cross-divisional plan for supporting curation, storage and dissemination of library-created or library managed digital content
  - build a robust, reliable, secure technical infrastructure base including both human and technology resources.
- **Consortia and contractual commitments:** UNTL maintains commitments to partner institutions of the UNT Libraries' Digital Collections, including The Portal to Texas History, as well as contractual agreements to assume or share in the responsibilities for preserving designated digital content.

5. Challenges

Recognized challenges in implementing an effective digital preservation program include, but are not limited to:

- **Rapid growth:** Technology that enables the preservation and dissemination of a variety of formats changes rapidly. As different types of materials are submitted (e.g., data sets, complex digital objects), monitoring different needs of the materials (storage size, metadata, etc.) and maintaining procedures and policies based on these needs is necessary.
- **Sustainability:** A sustainable digital preservation model should be developed that will respond to technological and staffing changes as needed, without under- or overestimating the needs imposed by these changes. The need for good cost models and affordable programs is widely acknowledged, yet still not fully addressed on a wider public scale. UNTL requires sufficient funding for operations and major improvements for digital asset management, as well as designated library funding to sustain ongoing preservation efforts. Further, there are administrative complexities in ensuring cost-effective and timely action to implement preservation strategies. The scale of funding is based on the level of commitment, therefore the program should reflect reasonable expectations of requisite resources, i.e., UNTL should not promise more than can be delivered.
- **Management:** Moving from well-managed digital collections to preserved collections requires institutional effort, partnership development, and a financial commitment. UNTL should provide a thoughtful balance between access and preservation, while being mindful of preservation's core role in maintaining access.
- **Partnerships:** UNTL must work with creators and providers of crucial content to employ appropriate maintenance prior to deposit that will facilitate future preservation.
- **Flexibility:** The digital preservation plan must continually review its abilities to respond to the evolving technological capabilities and changing user expectations without jeopardizing the ongoing care of the digital content.
- **Expertise:** UNTL must commit to continually updating staff expertise, where appropriate, as technologies change.
- **Rights:** There may be intellectual property and other rights-based constraints on providing access that impact digital preservation efforts.
- **Privacy:** As UNTL strives to make increasingly more digital collections openly available to the world, personal information might appear within these collections that violates the privacy of the people whom this information represents.

6. Principles

6.1 Guiding principles

UNTL will use consistent criteria for selection and preservation as with other resources in the libraries. Materials selected for digital stewardship and preservation carry with them UNTL's commitment to maintain the materials for as long as needed or desired.

- The Libraries are committed to the long term preservation of selected content.
- Digital preservation is an integral part of UNTL's process.
- Processes, policies, and the institutional commitment are transparently documented.
- Levels of preservation and time commitments are determined by selectors and curators, in consultation with technical experts.
- UNTL will participate in the development of digital preservation community standards, practices, and solutions.

6.2 Operating principles

The Libraries will strive to:
- Develop a scalable, reliable, sustainable, and auditable digital preservation infrastructure.
- Manage the hardware, software, and storage media components of the digital preservation function in accordance with environmental standards, quality control specifications, and security requirements.
- Comply with the Open Archival Information System (OAIS) reference model and other appropriate digital preservation standards and practices.
- Ensure that the digital archive is as interoperable as possible by utilizing open source options whenever feasible.
- Ensure the integrity of the data within the digital preservation infrastructure.
- Secure metadata (e.g., administrative, descriptive, preservation, provenance, rights, and technical information) necessary for access to and the use of the digital assets.
- Comply with copyright, intellectual property rights, and/or other legal rights related to copying, storage, modification, and use of digital resources.

6.3 Standards

- UNT is best served when distributed and disparate systems conform to standards and best practices that make communication between these storage systems possible.
- To utilize the OAIS Reference Model as the basis for developing and implementing strategies and tools for long-term digital information preservation and access.

7. Categories of Commitment

UNTL's levels of commitment as outlined below recognize that developing solutions for "born digital" materials informs solutions for the other categories; it does not imply that these assets are inherently more valuable or important than any of the other categories and/or our traditional, analog materials.

- **Born-digital materials:** Rigorous effort will be made to ensure preservation in perpetuity of materials selected for preservation, both library resources and institutional records.
- **Digitized materials (no available analog):** Every reasonable step will be taken to preserve materials without a print analog, when re-digitizing is not possible or analog versions are located elsewhere. Also included are digitized materials that have annotations or other value-added features making them difficult or impossible to recreate.
- **Digitized materials (available analog):** Reasonable measures will be taken to extend the life of the digital objects with a readily available print analog. However, the cost of re-digitizing as needed will be weighed against the cost of preserving the existing digital objects.
- **Commercially available digital resources:** UNTL has responsibility for working externally through consortia, licensing agreements, etc. to assure that one party or parties provides the necessary infrastructure to provide for preservation activities so that UNT faculty, staff, and students will have adequate ongoing access to commercially available digital resources. If the resources are external to UNTL, there needs to be an articulated exit strategy in the event of the cessation of the consortia or licensing agreements. Particular emphasis should be given to resources which exist in digital form only.
- **Other items and materials:** No preservation steps will be taken for materials requested for short-term use, such as materials scanned for E-reserve and document delivery, or for content that does not meet the requirements of the Collection Development Policy for the UNT Libraries' Digital Collections

8. Levels of Preservation

UNTL adheres to the National Digital Stewardship Alliance (NDSA) Levels of Digital Preservation developed by NDSA, as a tiered set of recommendations on how organizations should begin to build or enhance their digital preservation activities. to ensure secure, long-term integrity to its digital collections. Based on the current digital preservation standards, UNTL will apply the NDSA levels as a way of identifying gaps and improving its digital preservation infrastructure.

8.1 NDSA Levels of Preservation

Table 1: Version 1 of the Levels of Digital Preservation

	Level 1 (Protect your data)	Level 2 (Know your data)	Level 3 (Monitor your data)	Level 4 (Repair your data)
Storage and Geographic Location	- Two complete copies that are not collocated - For data on heterogeneous media (optical discs, hard drives, etc.) get the content off the medium and into your storage system	- At least three complete copies - At least one copy in a different geographic location - Document your storage system(s) and storage media and what you need to use them	- At least one copy in a geographic location with a different disaster threat - Obsolescence monitoring process for your storage system(s) and media	- At least three copies in geographic locations with different disaster threats - Have a comprehensive plan in place that will keep files and metadata on currently accessible media or systems
File Fixity and Data Integrity	- Check file fixity on ingest if it has been provided with the content - Create fixity info if it wasn't provided with the content	- Check fixity on all ingests - Use write-blockers when working with original media - Virus-check high risk content	- Check fixity of content at fixed intervals - Maintain logs of fixity info; supply audit on demand - Ability to detect corrupt data - Virus-check all content	- Check fixity of all content in response to specific events or activities - Ability to replace/repair corrupted data - Ensure no one person has write access to all copies
Information Security	- Identify who has read, write, move and delete authorization to individual files - Restrict who has those authorizations to individual files	- Document access restrictions for content	- Maintain logs of who performed what actions on files, including deletions and preservation actions	- Perform audit of logs
Metadata	- Inventory of content and its storage location - Ensure backup and non-collocation of inventory	- Store administrative metadata - Store transformative metadata and log events	- Store standard technical and descriptive metadata	- Store standard preservation metadata
File Formats	- When you can give input into the creation of digital files encourage use of a limited set of known open formats and codecs	- Inventory of file formats in use	- Monitor file format obsolescence issues	- Perform format migrations, emulation and similar activities as needed

National Digital Stewardship Alliance (NDSA) (2012). NDSA Levels of Preservation
    Retrieved from
    http://www.digitalpreservation.gov/ndsa/activities/levels.html

9. Roles and Responsibilities

UNTL has identified the following stakeholder categories for the digital preservation program. The terminology is adapted from the OAIS Reference Model (CCSDS 650.0-M-2 (2012)).

- **Producer:** is the role played by those persons or client systems that provide the information to be preserved. Producers include faculty, students, staff, alumni, collectors, creators of content, publishers, and others. Producers will be responsible for complying with established deposit requirements and working with the management of the digital archive to ensure a successful transfer. (expanded OAIS definition)
- **Management:** is the role played by those who set overall OAIS policy as one component in a broader policy domain, for example as part of a larger organization. UNTL's Deans' Council will be responsible for setting digital preservation policies and integrating them into broader organizational contexts. (expanded OAIS definition)
- **Administrators:** encompass content stewards (designated staff responsible for selection and for ongoing curation of specific collections), digital preservation specialists, and working teams (see appendix for list). Administrators will be responsible for the establishment of the digital preservation program and for day-to-day management of the digital archives. [Note: OAIS uses Administration Functional Entity: The OAIS functional entity that contains the services and functions needed to control the operation of the other OAIS functional entities on a day-to-day basis.]
- **Co-operating Archives:** includes those Archival Institutions that have Designated Communities with related interests. They may order and ingest data from each other. At a minimum, co-operating Archives must agree to support at least one common Submission Information Package (SIP) and Dissemination Information Package (DIP) for inter-Archive requests. At UNTL we think of this group as collaborators. Examples include: Texas Digital Library (TDL).
- **Consumer:** represents the role played by those persons, or client systems, who interact with OAIS services to find preserved information of interest and to access that information in detail. This can include other OAISes, as well as internal OAIS persons or systems.
- **User Groups / Client Groups:** include the various types of clients who use the UNT Libraries' Digital Collections

10. Collaboration/Cooperation

UNTL acknowledges digital preservation as a shared community responsibility, and as such has long-standing and emerging partnerships with similarly committed organizations (e.g., TDL and MetaArchive) and is committed to collaborating with other institutions, in addition to units internal to UNT in order to:

- Advance the development of the UNTL digital preservation program.
- Share lessons learned with other digital preservation programs.
- Extend the breadth of our available expertise.
- Extend the digital content that is available within a broad information community to UNTL users through cooperative efforts.

Generally, in working, cooperating, and collaborating with others, UNTL will strive to:

- Understand the goals, objectives, and needs of the communities of creators and the communities of consumers of its digital resources.
- Identify appropriate partners and stakeholders to contribute to national and international efforts in digital preservation .
- Help develop national and international strategies and initiatives that enable the distribution of collection, description, service delivery, digitization, and preservation activity.
- Work actively with creators of digital materials to encourage and promote standards and practices.

11. Access and Use Criteria

UNTL acquires, manages, and preserves digital resources so that they remain accessible to its constituents over the long term. Certain limitations may be placed on access due to legal, donor, and/or other restrictions; generally, to the extent that it is possible, UNTL endeavors to make its digital resources accessible to all users. Additional restrictions may apply to usage of items. See the UNT Libraries' Digital Collections Usage and Feedback Policy, sections 2. Access and 3. Rights/Terms of Use for complete guidelines.

12. Implementation

Implementation of this policy framework is contingent upon the infrastructure (technical and human resources) provided by UNT and UNTL, the availability of cost-effective solutions, the adoption of standards, and the evolution of best practices and procedures.

13. Review Cycle

This policy will be reviewed and updated as needed with a full review every two years to assure timely revisions as technology progresses and preservation strategies and experience mature.

Appendix 1: Glossary

Access: The process for the retrieval of data and information from storage media, through the use of catalogs, indexes, and/or other tools.

Acquire: To take physical and legal custody of data and information.

Analog: Data and information in a format that must be digitized to make it digitally accessible.

Aubrey: Framework developed by the UNT Libraries to provide end user access to digital collections in for the University of North Texas Libraries.

Bit-Level Preservation: Minimum digital preservation standard; the goal is to maintain the integrity of the original bit-stream of a digital object. It is accomplished by maintaining backup copies (onsite and/or offsite), the periodic refreshing of those copies to new storage media, and conducting fixity checking.

Born-Digital: Data and information created and maintained in a digital format.

Coda: Archival management system created by the UNT Libraries for registering, storing, replicating, and verifying fixity of preservation files.

Collection: A group of materials assembled by a person and/or organization, with one or more unifying characteristic.

Curation: Activities related to managing data and information throughout its lifecycle, ensuring that data are properly appraised, selected, and securely stored, while appropriately maintaining logical and physical integrity and authenticity. Further, that data is made and remains accessible and viable in subsequent technology environments.

Data Sets: Collections of data. The data formats include, but are not limited to, flat file tabular data, relational databases, text corpora, qualitative data in field notes, scholarly editions and thematic research collections.

Digital Assets: Digital objects (e.g., text, image, audio-visual files) owned or managed by an institution (or person).

Digital Collections: The collective digital library interfaces operated by the UNT Digital Libraries comprising The Portal to Texas History, the UNT Digital Library, and The Gateway to

Oklahoma History. When used in the capitalized form it is meant to refer to the UNT Libraries' Digital Collections.

Digital Object: An entity in which one or more content files and their corresponding metadata are united, physically and/or logically, through the use of a digital wrapper.

Digital Preservation: Comprehensive set of managed activities that are necessary to provide continued access to digital objects, beyond the limits of media failure or technology change. At minimum it should include bit-level preservation.

Digital Rights Agreement: A legal document that provides the UNT Libraries a non-exclusive license to preserve and provide access to resources owned by the rights owner for digital content in the UNT Libraries' Digital Collections.

Digital Wrapper: A technology that encapsulates administrative, technical, preservation, descriptive, or structural metadata and possibly content files into a specified format for serialization or transmission. An example of this is the Metadata Encoding & Transmission Standard (METS).

Digitized Materials: Analog materials that have been transformed into digital form, especially for storage, access and use in a computer environment.

EUID: "Enterprise UserID" is the local terminology for the unique identifier given to students, faculty and staff at the University of North Texas as part of the authentication framework in use by the university.

Institutional Records: Data or information in a fixed form, regardless of medium, that is created or received in the course of institutional activities and maintained as evidence of that activity for future reference.

Object Fixity: The quality of a digital object to be stable and resist change.

On-Campus: Physical UNT locations in Denton including the main campus and Discovery Park.

Open Archival Information System (OAIS) Reference Model: A theoretical framework that describes the components and processes necessary for a digital archive, including six distinct functional areas: ingest, archival storage, data management, administration, preservation planning and access. (A Glossary of Archival and Records Terminology, Society of American Archivists 2005) Full reference model and specifications: Reference Model for an Open Archival Information System (OAIS), Recommended Practice, CCSDS 650.0-M-2, Magenta Book, Space Communications and Navigation Office, NASA, June 2012

Partners: Person(s) or organization(s) that contribute materials to the UNT Libraries' Digital Collections, and who are usually the content owners or rights holders for the materials.

Partnership Agreement: A legal document that defines the relationship between the UNT Libraries and a contributing partner as well as the various rights each retains to the digital content in the UNT Libraries' Digital Collections.

Provenance: Information regarding the origins, custody, and ownership of an item or collection. (Richard Pearce-Moses, A Glossary of Archival and Records Terminology, Society of American Archivists 2005)

Reformatting: The process of creating a copy with a format or structure different from the original, for preservation and/or access; this may be accomplished via, transcription, xerography, microfilming, and/or digitization.

Scholarship: Use of intellectual resources, such as those managed by the UNT Libraries, for research, teaching, and learning.

Stewardship: The responsibility for overseeing and protecting materials and assets.

Sustainable Access: The process of providing long-term resources (fiscal, human, and technological) necessary to maintain access to information and digital objects in a repository.

Trusted Digital Repository: A repository whose mission is to provide reliable, long term access to managed digital resources to its designated community, now and in the future. "Trustworthiness" should be quantifiable via an assessment tool applied to a repository conceptual model (e.g. TRAC / ISO 16363 for for OAIS Reference Model).

University of North Texas Libraries (UNTL): The library managing the digital collections discussed throughout the Digital Preservation policy documentation.

UNT Community: Individuals and entities currently affiliated with the University of North Texas that have a valid/active EUID.

UNT Digital Libraries: Administrative unit in the UNT Libraries that manages digital and web-based systems.

UNT Extended Community: UNT students, faculty, staff, alumni, and administrators.

Appendix 2: Sources Consulted

The following resources were consulted in the development of this policy framework:

Cornell University Library *Digital Preservation Policy Framework* (December 2004)
Ohio State University *Digital Preservation Policy Framework* (August 2013)
Reference Model for an Open Archival Information System (OAIS), Recommended Practice,
        CCSDS 650.0-M-2 (Magenta Book) Issue 2, June 2012
Society of American Archivists *A Glossary of Archival and Records Terminology*, 2005