

# University of Houston Libraries

## Digital Preservation Policy

Updated February 2018

### Introduction

The following digital preservation policies are guided by the “Action Plan for Developing a Digital Preservation Program.” Conforming to the OAIS Reference Model and Trusted Digital Repository guidelines, the Action Plan walks institutions through the process of establishing a high-level framework, creating policies and procedures, building technological infrastructure, and addressing resources needed to sustain a digital preservation program for the long term.

UH Libraries Digital Preservation Policy consists of four main sections:

1. **Digital Preservation Policy:** provides a comprehensive overview of what is included in a digital preservation program, including: Purpose, Objectives, Mandate, Scope, Challenges, Principles, Roles and Responsibilities, Collaboration, Selection and Acquisition, and Access and Use.
2. **Appendix A: Policy Implementation:** describes the implementation at UHL of digital preservation policies, procedures, roles, and responsibilities, in greater detail than the broad-level policy framework.
3. **Appendix B: Technological Infrastructure:** outlines digital preservation system functions and requirements in greater detail than the policy framework.
4. **Appendix C: Digital Preservation Glossary:** a glossary of digital preservation terms used in this document and their definitions.

### Policy Review

Members of the Digital Preservation Working Group will review and update this document annually.

### Revisions to Date

- 2018-02-01: updated roles and responsibilities, added technical information on new SIP specifications, and added procedures for handling confidential or sensitive data.
- 2017-02-20: restructured overall policy document, changed implementation and technological infrastructure sections to appendices.
- 2016-12-21: updated the document review schedule, added changes in roles and responsibilities for administration and management of the digital preservation program, and added information on new workflows created for the DAMS Implementation project.

- 2015-07-28: migrated policy to UH Libraries Digital Initiatives Policies and Guidelines.
- 2015-09-16: approved by UHL Administration.
- 2015-05-13: completed by Digital Preservation Task Force.

# Digital Preservation Policy

## Purpose

University of Houston Libraries aims to anticipate the needs of students, faculty, and staff within the context of a nationally-competitive public research university environment. By serving as a trusted steward of digitized and born-digital content, UH Libraries enables long-term access to intellectual and cultural materials of both regional and national interest. The Digital Preservation Policy supports the missions of UH and UH Libraries and makes explicit UH Libraries' commitment to preserving the digital assets in its collections through the development and evolution of a comprehensive digital preservation program. The policy reflects the goals defined in our institutional mission and contains references to other relevant UH Libraries policies and procedures. The audience for this policy includes librarians and staff of UH Libraries, digital content donors/depositors, funders, and users.

### Related Documentation:

- [UH Mission Statement](#)
- [UH Libraries Mission Statement](#)

## Objectives

The Libraries' preservation role is guided by our responsibilities to manage and share our collections for local, national, and global access. By ensuring continued access to these valuable and unique resources, we are protecting substantial institutional investments and supporting the University's goal to establish itself as a preeminent public research university in the 21st century.

The Libraries defines the primary objective of digital preservation activities as maintaining the ability to meaningfully access digital collection content over time. The primary concern is preserving the ability to access the archival digital object from which derivative files may be created or re-created over time. To this end, preservation of digital library material includes:

- Maintaining multiple, geographically-distributed copies of all digital assets falling within the scope of the digital preservation policy.
- Strategically monitoring digital assets for format obsolescence and data integrity.
- Repairing and replacing corrupted data.
- Creating, storing, and maintaining ongoing preservation metadata.
- Performing format migrations and/or emulation as needed.
- Ensuring that authenticity and provenance is maintained.
- Understanding and reporting on risks affecting ongoing access.
- Periodically reviewing preferred formats and digital media standards.

- Ensuring our approach is based on the preservation functions of the OAIS model.
- Conducting audits to establish trustworthiness and evaluate workflow efficiency.

## Mandate

UH Libraries' digital preservation mandate is drawn from the following areas.

- **Organizational Commitment:** UH Libraries' digital preservation mandate is drawn from the charge of the Digital Preservation Task Force, formed in May of 2014, and the charge of the Digital Preservation Working Group, developed in November 2016. This statement establishes administrative support and provides direction for the development of a comprehensive digital preservation program aimed at maintaining long-term access to the Libraries' digital assets.
- **Scholarship:** As an institution of higher education, the University of Houston is obligated to support scholarship, teaching, and learning. As more resources and services associated with these functions become digital, UH's responsibilities must expand to include the identification, stewardship, and preservation of designated digital content.
- **Institutional Records:** UH Libraries' University Archives are charged with collecting and preserving university records that document the history of the University of Houston, including those in electronic format. University Archives records, increasingly born-digital, are frequently requested by units on campus to fulfill administrative and even legal needs.
- **Legal Obligations:** As a state and government institution, the University of Houston has legal obligations derived from federal and state laws such as the State Records Retention Schedule and the Freedom of Information Act that require us to maintain the authenticity and integrity of archival files.
- **Consortia and Contractual Commitments:** UH Libraries has commitments to TDL and contractual agreements in association with our preservation efforts.

### Related Documentation:

- [UH Libraries' Digital Preservation Working Group Charge](#)

## Scope

As a result of our institutional mandate, UH Libraries has responsibility for the systematic management, storage, preservation, and accessibility of its digital assets. Taking into consideration that the fast pace of technological change greatly impacts digital preservation plans and techniques, UH Libraries has prepared this policy to serve as a guide for moving forward into the immediate future and beyond.

UH Libraries recognizes that the following assets will be retained and managed in accordance with this preservation policy:

- Digital versions of resources owned and reformatted by UH Libraries and that fall under the parameters of UH Libraries' Digital Collection Development Policy.

- Unique born-digital resources that are part of UH Libraries' archival/manuscript collections and which are unlikely to be preserved anywhere else.
- Digitized master files of extremely high-risk audiovisual content, representing a significant financial and institutional investment.
- Any other content acquired or digitized by UH Libraries that falls under the parameters of UH Libraries' Digital Collection Development Policy.

Content acquired exclusively by repositories which are governed by partner organizations (Including the HathiTrust Digital Library and TDL) will be deferred to those organizations for ongoing preservation and/or management. This content includes electronic theses and dissertations (ETDs).

The preservation and retention policy will be reviewed by the end of each calendar year to assure timely updates as technology and experience mature, or more often if the need arises.

**Related Documentation:**

- [UH Libraries' Digital Collection Development Policy](#) (PDF)

## Challenges

As with any digital preservation program, UH Libraries faces multiple challenges and risks.

Challenges include:

- The development of organizational and technical infrastructures which can efficiently manage digital content is a complex and expensive undertaking.
- Maintaining understandable, reliable, and authentic digital materials and ensuring their accessibility over the long term requires the capacity for active management of rapidly changing hardware and software. A method for managing persistent identifiers for digital materials must be put into place, and must be sustainable over time. Currently, persistent identifiers are created and managed via the Greens ARK minter app within the larger digital access and preservation workflow.
- A wide variety of variables exist during the digital preservation process: mixed levels of complexity in terms of object structure, relationships, and dependencies; mixed levels of intellectual control; different levels of complexity in preservation planning and processing; different timetables for preservation action; and the need of different preservation approaches, often at different scales.
- A tremendous amount of digital material of all types is being produced and an exponential growth is occurring.
- Digital assets will need to be migrated as systems change over time. Standards which will aid in the preservation of digital materials are not yet well established, and this lack of consistency inhibits concerted preservation efforts.

# Principles

To fulfill our institutional mandate, UH Libraries adheres to a series of principles that support long-term access to materials. Some of these principles, including continual improvement, excellence in collections, partnership and collaboration, and service excellence, are rooted in UH Libraries core values as articulated in the [UH Libraries 2017-2021 Strategic Directions](#) document. The principles demonstrate the central role digital preservation plays in the success of UH Libraries, and aligns UH Libraries with important national standards and best practices. The following elements of the digital preservation framework allow us to uphold the principles of the Libraries.

## ***Documentation***

- Clearly and consistently document policies, procedures, and practices and review them on a regular basis.
- Document and identify any remedial treatment, alterations, and/or additions to master files and make them as minimal and reversible as possible.

## ***Management***

- Manage files throughout their lifecycle by determining their retention, use, and preservation, ideally at acquisition or creation.
- Comply with intellectual property, copyright, and ownership rights related to digital assets through the use of a defined [copyright assessment protocol](#) created by UH Libraries' RightsStatements.org working group. This document is evolving alongside the development of our digital asset management system.
- Establish procedures to meet archival requirements pertaining to provenance, chain of custody, authenticity, and integrity.
- Select the most appropriate and cost-effective strategy for the preservation of digital assets.
- Provide the widest possible access to digital objects selected for preservation.

## ***Metadata***

- Generate preservation metadata necessary for the future use of digital assets.
- Preserve descriptive metadata and maintain the links between digital assets and their metadata. This will be achieved via the inclusion of the preservation package ARK identifier in the item-level descriptive metadata for each access object.
- Preserve digital assets that have been described using appropriate metadata for resource discovery, management, and/or preservation.
- Use standardized metadata schema for interoperability between programs.

## ***Standards***

- Ensure our approach is based on the OAIS Model and the requirements for a Trusted Digital Repository.

- Adhere to prevailing community standards for preserving access to digital assets.
- Commit to an interoperable, scalable digital archive with appropriate storage management capabilities.

### ***Technology***

- Manage the hardware, software, and storage media components of the digital preservation function in accordance with environmental standards, quality control specifications, and security requirements.
- Keep abreast of technological change and monitor new developments in the digital preservation field, including an awareness of international standards.

### **Related Documentation:**

- [Trusted Digital Repositories: Attributes and Responsibilities](#) (PDF)
- [Reference Model for an Open Archival Information System](#) (PDF)

## **Roles and Responsibilities**

UH Libraries has identified multiple stakeholders who contribute to the life cycle of records and administer the digital preservation program. These core groups include UH Libraries as an institution, members of the UH Libraries' administration, Library Technology Services, Metadata and Digitization Services, Special Collections, Branch Libraries, Digital Research Services, the larger UH Main Campus community, TDL, and content producers and donors. This section articulates the key roles of each group.

### ***UH Libraries***

Since its beginning, the UH Libraries has been entrusted to preserve important scholarly resources in order to fulfill the educational and administrative mission of the University. UH Libraries continues this commitment with the digitized and born-digital assets it retains.

### ***Administration***

Digital assets retained, managed, or owned by UH Libraries need adequate resources and support to remain accessible into the future. UH Libraries' administration allocates appropriate resources to ensure that preservation activities are sustained and developed into the future.

### ***Library Technology Services***

The role of technology is an essential leg of a robust and sustained digital preservation program. The program relies on Library Technology Services to support key software applications and maintain local and remote storage so that it aligns with UH Libraries' digital preservation principles as well as digital preservation national standards and best practices.

### ***Metadata and Digitization Services***

A successful digital preservation program requires designated librarians to be engaged with the work of digital preservation on a day-to-day basis. Members of Metadata and Digitization

Services manage key portions of the digital preservation program as one part of a larger digital library workflow.

### ***Special Collections***

As the individuals who have specialized content knowledge and established relationships with content producers, librarians and archivists in Special Collections play a key role in identifying, acquiring, and describing content for the digital preservation program. Additionally, designated archivists adhere to the digital preservation program when managing born-digital workflows.

### ***Branch Libraries***

As the individuals who have specialized content knowledge and established relationships with content producers, librarians in the Architecture and Art, Music, and Health Sciences Libraries play a key role in identifying, acquiring, and describing content for the digital preservation program.

### ***Digital Research Services***

Digital Research Services coordinates with other internal and external stakeholders to develop and implement digital preservation policies and to ensure the policy and strategies are being met, especially for resources managed by or through external stakeholders.

### ***UH Main Campus Community***

Campus units deposit materials of enduring value with the Libraries to be preserved. The Libraries receives UH funding that it chooses to allocate towards digital preservation.

### ***Texas Digital Library***

TDL manages and supports long-term, distributed storage for UH Libraries' digital preservation program.

### ***Content Producers and Donors***

Key stakeholders include the producers or creators of the digital content who donate/deposit digital assets with UH Libraries with the understanding that their gifts will be properly managed and preserved.

For detailed information on digital preservation roles and responsibilities at UH Libraries, see Appendix A, Section 1: [“Roles and Responsibilities.”](#)

## **Collaboration**

As digitized and born-digital assets grow, the ability for any one institution to preserve digital assets without collaboration involving other groups and institutions becomes increasingly difficult. To address increasing digital preservation needs, UH Libraries aims to collaborate with external partners, including TDL and HathiTrust, to implement the digital preservation policy and

to share the larger responsibility of preserving digital assets for long-term access. In doing so, UH Libraries will take appropriate steps to build and sustain these relationships, including:

- Identifying appropriate partners.
- Establishing written agreements on roles and responsibilities.
- Helping develop the policies, procedures, and tools necessary to support any collaborative digital preservation efforts
- Working with the producers of digital assets to encourage practices that enable preservation.
- Sharing information on its own experiences, and learning from the experience of others.

## Selection and Acquisition

The digital preservation program supports existing digitization and born-digital workflows at UH Libraries. As such, the program complies with current policies and procedures around the selection and acquisition of content for long-term access, including:

- [UH Libraries' Digital Library Collection Development Policy](#) (PDF)
- [Special Collections' Procedures for Accessioning Born Digital Content](#)

For detailed information on digital preservation selection and acquisition at UH Libraries, see Appendix A, Section 2.b: ["Selection and Acquisition Policies and Procedures."](#)

## Access and Use

UH Libraries' aim is to provide online access to all digital collections in a meaningful format for its users via the Libraries' Web site and other dissemination platforms. To facilitate this process, UH Libraries collects and communicates relevant information about rights and access restrictions to its users. Information on the copyright status of materials will be conveyed in the UH Digital Library using the standardized statements provided by [RightsStatements.org](#); work to update descriptive metadata with these statements is underway as of December 2016. Access restrictions might relate to intellectual property rights, legal requirements, or privacy concerns.

For detailed information on digital access and use at UH Libraries, see Appendix A, Section 2.d: ["Access and Use Policies"](#) and Appendix B, Section 3.h: ["Access."](#)

## Review Cycle

This policy is reviewed annually by the Digital Preservation Working Group to accommodate changes in institutional resources, technology, and priorities.

## References

Columbia University, 2006

<http://library.columbia.edu/services/preservation/dlpolicy.html>

ICPSR Digital Preservation Policy Framework, 2007

<http://www.icpsr.umich.edu/icpsrweb/content/datamanagement/preservation/policies/dpp-framework.html>

Government of Canada, Digital Preservation Policy Framework: Development Guideline

<http://canada.pch.gc.ca/eng/1443189702298>

National Library of Australia, 2013

<http://www.nla.gov.au/policy/digpres.html>

NDSA Levels of Digital Preservation, 2013

[http://digitalpreservation.gov/ndsaworkinggroups/documents/NDSA\\_Levels\\_Archiving\\_2013.pdf](http://digitalpreservation.gov/ndsaworkinggroups/documents/NDSA_Levels_Archiving_2013.pdf)

The Ohio State University Libraries, 2013

[http://library.osu.edu/documents/SDIWG/Digital\\_Preservation\\_Policy\\_Framework.pdf](http://library.osu.edu/documents/SDIWG/Digital_Preservation_Policy_Framework.pdf)

Purdue University Research Repository Digital Preservation Policy, 2012

<https://purr.purdue.edu/legal/digitalpreservation>

The State and University Library, Denmark, 2012

<http://en.statsbiblioteket.dk/about-the-library/dpstrategi>

University of Massachusetts, Amherst Libraries, 2011

<http://www.library.umass.edu/assets/aboutus/attachments/University-of-Massachusetts-Amherst-Libraries-Digital-Preservation-Policy3-18-2011-templated.pdf>

YUL Digital Preservation Policy Framework, November 2014

<http://www.library.yale.edu/iac/DPC/revpolicy2-19-07.pdf>

# Appendix A: Policy Implementation

## 1. Roles and Responsibilities

**Creators/Producers:** The role played by those persons or client systems that provide the information to be preserved. Creators/producers include faculty, students, staff, alumni, collectors, content creators, publishers, and others. Creators/producers can also be internal persons or systems. They can generate “born-digital” content or digitized surrogates from physical objects. Creators/producers will be responsible for complying with established deposit requirements and working with the management of the UH Libraries digital preservation program to ensure a successful transfer.

**Management:** The chair of the Digital Preservation Working Group, working with other key stakeholders, will be responsible for setting digital preservation policies and integrating them into broader organizational contexts.

**Digital Preservation Administrators:** Designated staff responsible for selection and for ongoing curation of specific collections (see below). Digital Preservation Administrators will be responsible for the establishment and day-to-day management of the digital preservation program.

**Cooperating Repositories:** Those repositories that have designated communities with related interests. They may ingest and provide access to each other’s collections. At a minimum, cooperating repositories must agree to support at least one common Submission Information Package (SIP) and Dissemination Information Package (DIP) among repositories. Examples include: UH Digital Library (UHDL), UH Libraries’ Institutional Repository (through TDL partnership), DuraSpace (through TDL partnership), the Texas Data Repository (through TDL partnership), and HathiTrust Digital Library.

**Consumers/Client Groups:** The role played by those persons or client systems who interact with UH Libraries repositories to access information of interest. This can include other institutions and repositories, as well as internal persons or systems.

**Chart 1: Digital Preservation Management and Administrators**

Position	Individual	Responsibilities
Head of Digital Research Services	Santi Thompson	Coordinates with internal and external partners (Including TDL and HathiTrust) to implement digital preservation policy and program; collaborates with the Head of Library Technology Services, Digital

		Projects Coordinator, and the Digitization Services Coordinator to conduct audit
Head of Library Technology Services	Rob Spragg	Collaborates with DPWG to implement digital preservation policies and procedures; maintains hardware and software for digital preservation program including distributed digital preservation service; provides support for Archivematica; assists with digital preservation audit process
Head of Metadata and Digitization Services	Annie Wu	Collaborates with DPWG to implement digital preservation policies and procedures; oversees digital preservation work done by librarians and staff members in MDS
Digital Projects Coordinator	Bethany Scott	Ensures digitized and born-digital content meets SIP requirements; creates AIPs for born-digital Special Collections content following digital preservation requirements; provides access to born-digital content; assists with digital preservation audit process
Systems Administrator 3	Marcus Elizondo	Provides server and storage support and error troubleshooting for Archivematica

## 2. Digital Assets

### 2.a Quality Creation and Benchmarking

#### **Image capture specifications**

UH Libraries bases its image capture specifications on the Federal Agencies Digitization Guidelines Initiative's (FADGI's) document [Technical Guidelines for Digitizing Cultural Heritage Materials](#) (PDF).

Technicians must capture each aspect of an object (page, recto/verso of photograph, etc.) as a 16-bit [TIFF](#) file, with the Adobe RGB 1998 colorspace embedded. All captures should be taken at 300 ppi or higher, depending on equipment used. Pixel dimensions vary based on the size of the original object. The TIFF file serves as the preservation master file.

For access files, technicians create 8-bit [JPEGs](#) with an embedded sRGB colorspace. Typically, the dimensions of JPEGs will not exceed 3,000 x 3,000 pixels.

For further details, see [FADGI Guidelines for Local Implementation](#) (PDF).

### ***Audio capture specifications***

UH Libraries bases its audio capture specifications on [Sound Directions: Best Practices for Audio Preservation](#) and the International Association of Sound and Audiovisual Archives' [Guidelines on the Production and Preservation of Digital Audio Objects](#).

All audio digitized by UH Libraries must be captured as 96 kHz, 24-bit [WAVE](#) files. Basic metadata is embedded in the file using FADGI's BWF Metaedit tool.

For access files, mid-range quality (44.1 kHz Stereo, 16-bit, 192 kbps, CBR) [MP3s](#) are created from the WAVE files. This process may include stitching files, increasing volume, and improving the clarity of sound.

### ***Video capture specifications***

Following such institutions as the Library of Congress, Library and Archives Canada, and 20th Century Fox, for film and video digitized by an outside vendor, we request a [lossless JPEG2000 in an MXF wrapper \(OP1a\)](#). This file serves as the preservation master. For video digitized by UH Libraries, the preservation master will be captured as an uncompressed AVI or MOV. This model was selected to expedite access to A/V holdings. As we expand in-house A/V digitization, we hope to move to a more robust software and hardware setup that would allow for the creation of MXF.

A [DV](#) created from the preservation master is used as a mezzanine/production-level file which is then edited and optimized for access and processed out as an [MP4](#).

Mezzanine files will not be preserved at this time due to limitations in archival storage; they will be stored on local network drives.

### ***Metadata specifications***

In order to keep master files accessible over time, and to ensure a digital object's authenticity and reliability, metadata must be collected to meet certain functional requirements. All content entering the preservation system must have descriptive, structural, and administrative metadata, and the metadata must be made available in well-documented and widely-adopted formats.

Sufficient metadata must be created to support a number of essential functions, listed below.

#### **1. Functions required of all digital masters**

- a. It will be possible to produce, in print or as an online (on-screen) display, a faithful, citable rendering of the physical source including the sequencing of its component parts (pages, volumes, boxes, folders, etc.)
- b. It will be possible to navigate sequentially through the physical components (go to next, previous, first, last, or nth page, etc.)
- c. The relationship between component parts of the physical source (pages, volumes, boxes, folders, etc.) will be represented.

- d. Images of blank pages, photographic versos, and other like materials will be included as sequenced components.
- e. It will be possible to associate higher-level descriptive metadata with digital component parts of the object (for example, for the purposes of citation.)

## **2. Functions required where applicable**

- a. Where possible, masters will support navigation to, between, and among logical structures (chapters, volumes, parts, boxes, folders). Citation of those features will also be supported.
- b. Where applicable and in a manner appropriate for the physical object in question, any enumeration found on the physical object will be represented. Representation will maintain all variations in the enumeration of the physical object's component parts (signature pages, preface, etc.)
- c. Placeholders for known missing materials (pages, photos, etc.) will be included as sequenced components. In the interest of creating complete digital masters, missing pages and other components should be identified as such in higher-level metadata. Where parts of the object are provided by third parties, information to that effect should be noted in descriptive metadata.

## **3. Functions strongly preferred**

- a. High-level logical structures will be identified in ways to enable more complex rendering and navigation functions. This would play an important role should objects become displayed in more associative, non-hierarchical contexts, such as image clouds or image recommendation services.
- b. For purposes of citation, access, etc., it will be possible to support association of higher-level metadata with both logical structures and specific instances of an object's representation.

### ***PREMIS implementation***

PREMIS metadata will be captured and stored within the canonical metadata documents associated with all digital objects. These values will be monitored on a regular basis to ensure fixity and ongoing access. For more information on what PREMIS semantic units or components are captured, refer to [Archivematica's PREMIS documentation](#).

## **2.b Selection and Acquisition Policies and Procedures**

The UH Libraries Digital Preservation Program acquires and preserves digital assets from the following sources: the UH Digital Library, UH Libraries' Institutional Repository, and digital archival materials from Special Collections. As a general operating procedure, content that is either digitized by, acquired by, or submitted directly to UH Libraries will be preserved. Content submitted by a producer to partner organizations, including TDL, is preserved by those organizations.

### ***Selection of digitized materials for UH Libraries' digital collections***

Selection criteria for objects deposited into digital repositories at UH Libraries should conform to the Digital Collections Development Policy, administered by the UH Libraries' Digital Collections Management Committee (DCMC).

While repositories may have specific selection criteria based on their scope and purpose, any UH Libraries repository should retain collections and items that generate national recognition for the University of Houston and UH Libraries. Additionally, collections and items should:

- Be of significant research and/or teaching interest
- Align with University of Houston strategic priority areas (Energy, Health, Arts)
- Exist nowhere else as digital content that is easily accessible and/or of comparable quality
- Meet existing or anticipated demand

Specific selection criteria for each repository can be found in [UH Libraries' Digital Collections Development Policy](#) (PDF).

### ***Selection of born-digital assets to be preserved in UH Digital Preservation Program***

The selection of materials in born-digital form for preservation is guided by the same general objectives that direct the selection of materials in other media. These collections include unique born-digital resources that are part of UH Libraries' archival/manuscript collections and which are unlikely to be preserved anywhere else.

### ***Note on selection of electronic theses and dissertations (ETDs)***

All ETDs produced for University of Houston masters and doctoral programs are submitted by students to the Vireo Thesis and Management System, hosted by TDL. ETDs without embargoes, or with expired embargoes, are ingested into and made available through the UH Libraries' Institutional Repository.

TDL stores backups of Vireo and the UH Libraries' Institutional Repository to Amazon in differing intervals (last five days, last day of month, last day of year.) These backups are also stored in Amazon's S3 preservation service, a cloud-based distributed digital preservation network.

See also Content Selection Strategies for Cloud-Based Storage in Appendix B, Section 3.g: ["Preservation Planning."](#)

### **Related Documentation:**

- [Amazon Glacier service](#)

### **Acquiring and deaccessioning digital assets**

Files are acquired from creators through methods that assure the authority and integrity of the files. UH Special Collections has outlined their procedures in [Special Collections Procedures for Accessioning Born Digital Content](#).

Digital assets that no longer support the teaching and research activities of the University of Houston, the scholarly community, and the general public may be deaccessioned from collections and will no longer be maintained or preserved by UH Libraries. The policies and procedures for this process will be determined by DCMC.

### **Format specifications for producers of born-digital content**

These recommendations are designed to serve as a general guideline for file formats for producers of born-digital content. The digital preservation workflow accepts all file formats; however, some formats are more sustainable and easier to preserve long-term. By contrast, it may not be possible to fully preserve higher risk formats over time, as hardware and software needed to read the files may become obsolete and normalization and migration lead to data loss. The UH Libraries Digital Preservation Program ranks a format as either high, moderate, or low preference based on amount of support required to maintain the file format and probability of long-term stability.

*High Preference* indicates that the formats have the most support and the highest probability of long-term stability. The formats are typically openly documented and not compressed (or have lossless data compression).

*Moderate Preference* formats do not meet the minimum requirements for long-term retention but come close, and for practical reasons may be necessary for long-term maintenance. These formats are more likely to require migration in order to remain renderable.

*Low Preference* formats are not recommended or supported for long-term preservation. These files may be difficult or even impossible to render or provide access to in the future. These formats are likely candidates for normalization or migration. Information is often lost during this process.

<b>Chart 2: File Format Preferences for Digital Preservation</b>			
<b>Content type</b>	<b>High Preference</b>	<b>Moderate Preference</b>	<b>Low Preference</b>
Word processing	PDF/A-1a (.pdf), OpenDocument Text (.odt)	PDF (.pdf), Microsoft Word (.doc), Microsoft Open XML (.docx), Rich Text Format (.rtf)	Corel WordPerfect (.wpd), Lotus WordPro (.lwp)

Plain text	Plain text (.txt), comma-separated file (.csv), tab-delimited file (.txt)		
Structural markup text documents	SGML with DTD/Schema, XML (.xml) with DTD/Schema		SGML without DTD/Schema, XML without DTD/Schema
Spreadsheets	OpenDocument Spreadsheet (.ods), comma-separated file (.csv), tab-delimited file (.txt), PDF/A-1a (.pdf)	Microsoft Excel (.xls), Microsoft Excel Open XML (.xlsx)	
Audio	WAVE format (.wav)	AIFF uncompressed (.aif, .aiff), standard MIDI (.mid, .midi), Windows Media Audio (.wma), MPEG3 (.mp3), MP2 AAC (.m4a)	Audio CD, DVD-Audio, QuickTime MP4 AAC Protected (.m4p, .m4b), QuickTime MP3, iTunes, RealAudio (.rm, .ra), Shorten (.shn), RIFF-RMID (.rmi), Extended MIDI (.xmi), Module Music Formats, Mods (.mod)
Video	Lossless JPEG2000 in and MXF wrapper (OP1a)	QuickTime (.mov), AVI (.avi), MPEG-1 (.mpg), MPEG-2 (.mpg), MPEG-4 (.mp4)	Windows Media Video (.wmv)
Images	TIFF (.tif, .tiff)	JPEG (.jpg, .jpeg), JPEG2000 (.jp2), PNG (.png), PDF/A-1a (.pdf), GIF (.gif)	RAW (.raw, various), Adobe Photoshop (.psd), Kodak PhotoCD, Encapsulated PostScript (.eps), FlashPix (.fpx), PDF (.pdf)

## 2.c Transfer Requirements and Deposit Guidelines

At UH Libraries, digital assets that are candidates for preservation originate in Metadata and Digitization Services (MDS) for digitized assets and in UH Special Collections for born-digital assets.

UH Libraries' digitization projects are initiated by a project plan meeting, bringing together MDS with relevant stakeholders, which would include Special Collections, the Architecture and Art Library, or the Music Library, to discuss project scope and parameters. Upon the completion of the project, Digital Preservation Administrators prepare this content for transfer to UH Libraries' digital preservation system.

With born-digital items, Special Collections negotiates an agreement with the information producer, which includes a deed of gift or deposit agreement. Archives personnel work with information producers to collect information detailing provenance, file organization, hardware and software needed to read files, and context of the files and how they were created. Once content is accessioned, Digital Preservation Administrators transfer it to UH Libraries' digital preservation system.

Digital Preservation Administrators will be responsible for ensuring that required descriptive, structural, and technical metadata is preserved. In preparing materials for transfer, both MDS and Special Collections create submission information packages (SIPs) that contain the materials necessary for the long-term preservation of digital assets. To ensure consistent quality across all submissions, the Digital Preservation Working Group has created SIP specifications aimed at preserving the digital assets themselves; the minimal descriptive metadata attributed to these assets that is necessary to ensure appropriate levels of reliability, authenticity, and provenance; and the structural metadata necessary to preserve a canonical record of original order and/or hierarchy (if applicable).

See Appendix B, Section 3.a: [“Pre-Ingest: Producer-Archive Interactions”](#) and Appendix B, Section 3.c: [“Ingest.”](#) Preparation of Digital Assets for Transfer, UHL Digital Preservation System SIP Specification, for details on these processes.

## 2.d Access and Use Policies

University of Houston Libraries acquires, manages, and preserves digital assets so that they remain accessible to its constituents over the long term. Certain limitations may be placed on access due to legal, donor, or other restrictions, but in general, insofar as possible, UH Libraries endeavors to make its digital assets accessible to all users.

Each individual digital collection will have its own defined restrictions for access and use. These restrictions may be determined by intellectual property rights, legal requirements, privacy

concerns, or a project's mission. UH Libraries provides access to its digital assets in such a way that all license and donor agreements are respected.

A preservation copy (or copies) of digital assets are kept in preservation storage, which prohibits direct public access. Public access, where appropriate, to derived copies of digital assets in the UHDL is provided through the Libraries' digital asset management system (DAMS) and the Libraries' Web sites. Dissemination of these digital objects is managed by the Metadata Unit through the DAMS/digital project workflow. External requests for master copies of digital assets will be reviewed on a case by case basis. Additionally, consumers are permitted to download certain static image files through the UHDL Digital Cart service.

Currently, Special Collections provides public access to processed born-digital files in the reading room. Future plans include expanding access through the same avenues as other UH digital materials.

For additional information, see Appendix B, Section 3.h: ["Access."](#)

### 3. Digital Preservation Strategies

In general, the preservation strategies utilized by UH Libraries are based on both the OAIS conceptual model and the Trusted Digital Repository specifications. All decision-making surrounding digital preservation stems, whenever possible, from this common core. To that end, UH Libraries will:

- Establish and maintain a robust preservation system that is able to ensure the reliability, authenticity, and provenance of digital objects. This repository is modular in its structure, so services can be added or updated over time as the information landscape evolves and preservation needs change.
- Manage risk through the Digital Preservation Working Group and its administrators, who are actively involved in the day-to-day preservation of digital content. Decisions regarding specific preservation strategies (format-specific, etc.) will be documented by members of the team and consolidated into standard preservation rules and operational procedures. Members will monitor research/developments in the field of digital preservation and will reevaluate existing documentation and procedures to ensure continued relevance.
- Educate all staff working directly with digital content so that sustainability and preservation become consistently taken into consideration throughout the digital object lifecycle. Care will be taken to ensure digital content falling within the scope of the repository will be prepared according to stringent submission specifications.
- Commit to budgeting for the long-term preservation of digital assets.
- Collaborate with other institutions and organizations to strengthen our commitment to digital preservation and to share resources and best practices.

For more information on digital preservation strategies, see Appendix B, Section 3.g: [“Preservation Planning.”](#) Strategic Priorities for Digital Preservation, 2015-2018.

**Related Documentation:**

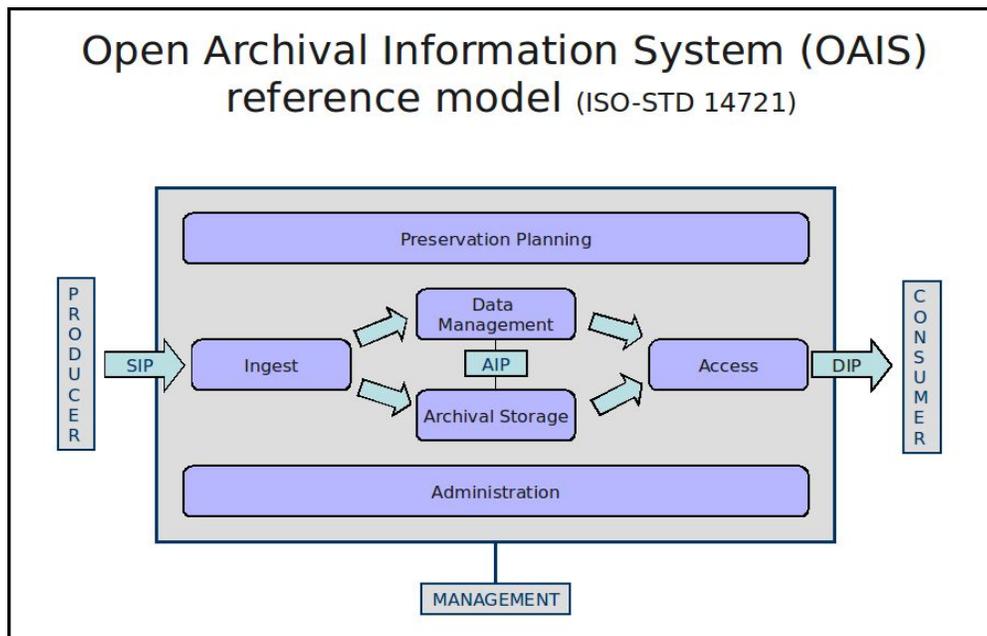
- [Trusted Digital Repositories: Attributes and Responsibilities](#) (PDF)
- [Reference Model for an Open Archival Information System](#) (PDF)

## 4. Technological Infrastructure

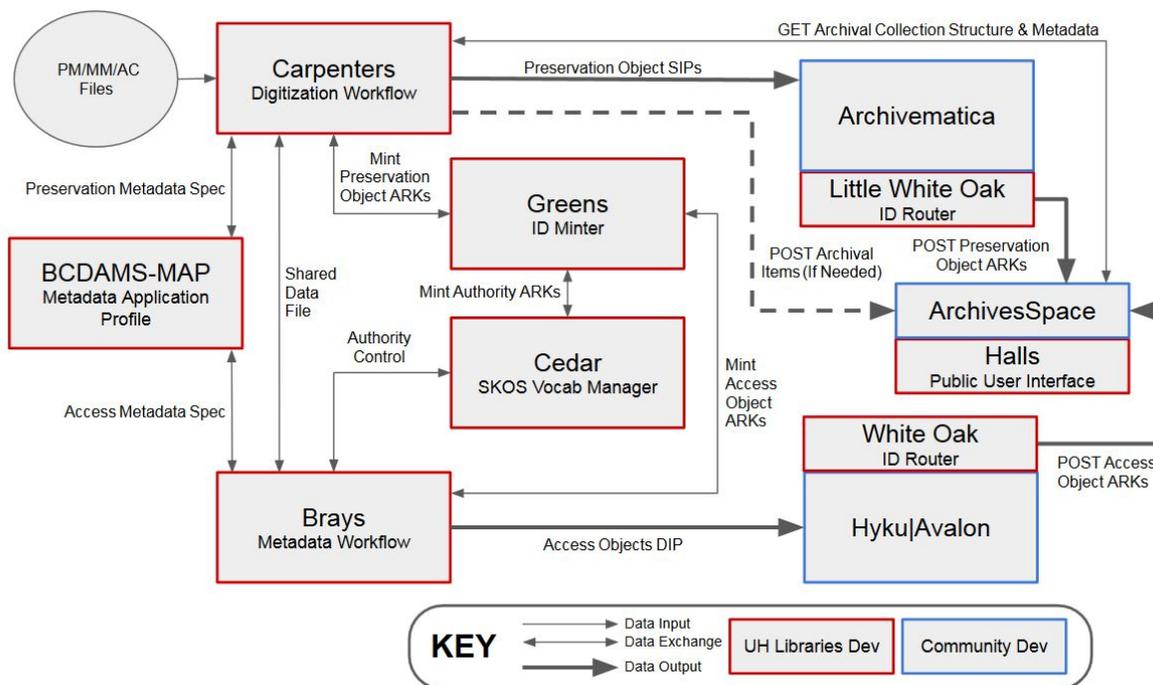
### 4.a Digital Archive Operations

The UH Libraries digital preservation workflow must process digital objects from ingest to archival storage and access in compliance with the ISO-OAIS functional model (Figure 1) and other digital preservation standards and best practices. The digital preservation function, including Archivematica, is situated within the larger digital access and preservation framework (Figure 2).

**Figure 1. OAIS Reference Model**



**Figure 2. UH Digital Access and Preservation Workflow**



#### 4.b Platform Requirements and Procedures

To meet current and future needs, the UH Libraries Digital Preservation System should fulfill the following requirements:

- **Scalability:** The ability for the repository to scale to manage large collections of digital objects.
- **Extensibility:** The ability to integrate external tools with the repository to extend the functionality of the repository, via provided software interfaces (APIs), or by modifying the code-base (open source software).
- **Interoperability:** The ability for the repository to interoperate with other repositories.
- **Security:** The system must provide multiple mechanisms to prevent compromise and data loss.
- **Performance:** The system must perform at a level that satisfies the needs of the users of system. It must have quick response times and high system availability.
- **Flexibility:** The ability for multiple instances for offsite recovery; the ability to function with the offsite backup facility; the ability for components to reside at different physical locations; the ability for development, testing, and production environments; capability for disaster recovery.
- **System support:** The quality of documentation and responsiveness of support staff or developer/user community (open source) to assist with problems.

- **Development community:** Reliability and support track record of the company providing the software; or size, productivity, and cohesion of the open source developer community.
- **Development organization:** Viability of the company providing the software; or stability of the funding sources and organizations developing open source software.
- **Technology roadmap for the future:** Technology roadmap that defines a system evolution path incorporating innovations and “next practices” that are likely to deliver value.

## References

Purdue University Libraries, “File Format Recommendations”  
<http://www.lib.purdue.edu/spcol/file-format-recommendations>

Digital Library Foundation, “Benchmark for Faithful Digital Reproductions of Monographs and Serials,” 2002  
<http://old.diglib.org/standards/bmarkfin.pdf>

# Appendix B: Technological Infrastructure

## 1. OAIS Reference Model

UH Libraries uses the [Open Archival Information System reference model \(OAIS\)](#) (PDF) as a basis and a means for communicating about digital preservation. The OAIS reference model is an ISO standard that outlines a common set of requirements and practices developed around the globe. OAIS also provides a common language for communication, collaboration, and sharing information across repositories. UH Libraries is committed to establishing and maintaining a digital preservation system that is OAIS compliant. Functional entities implemented in our system, such as pre-ingest, ingest, archival storage, data management, administration, preservation planning, and access are OAIS compliant. UH Libraries is establishing its program using a planning document designed around the OAIS model and the Trusted Digital Repository requirements and will periodically review our program to ensure ongoing compliance.

### Related Documentation:

- Brian F. Lavoie, [The Open Archival Information System Reference Model: Introductory Guide](#) (PDF) (Digital Preservation Coalition Technology Watch Report 04-01), January 2004

## 2. Information Packages

A basic concept of the OAIS Reference Model is the need to combine data and representation information into information objects. This model is valid for all the types of information in an OAIS compliant repository, such as the UH Libraries' digital preservation system. The OAIS Model contains the following three types of information packages:

*Submission Information Packages* (SIPs) are delivered by the producer to the digital preservation repository for use in the construction of one or more AIPs. Policies and procedures related to the delivery of digital content by the producer to the repository are further detailed in Appendix B, Section 3.a: [“Pre-ingest: Producer-Archive Interaction”](#) and Appendix B, Section 3.c: [“Ingest.”](#)

*Archival Information Packages* (AIPs) consist of the content information and the associated preservation description information (PDI), which is preserved within the digital preservation repository. See Appendix B, Section 3.d: [“Archival Storage.”](#)

*Dissemination Information Packages* (DIPs) are derived from one or more AIPs and received by the consumer in response to a request to the digital preservation repository. See Appendix B,

Section 3.h: [“Access”](#) for information on how access is provided to the consumer from the digital preservation repository.

**Related Documentation:**

- [Framework for the Definition of Significant Properties](#) (PDF), 2009, see 1.4 Significant Properties and the OAIS
- Library of Congress, [Archival Information Package Design Study](#) (PDF), 2001
- [FCLA Digital Archive \(FDA\) SIP Specification](#) (PDF), 2006

## 3. Functional Entities

### 3.a Pre-ingest: Producer-Archive Interaction

Digitization projects conducted by UH Libraries are initiated by a project plan meeting, which brings relevant stakeholders together, including as appropriate Special Collections, MDS, the Architecture and Art Library, the Music Library, and/or the Health Sciences Library, to discuss project scope and parameters. In addition to identifying physical condition concerns of analog materials and establishing a project timeline, the meetings outline instructions for digital object creation, including digitization and metadata specifications. Upon the completion of the project and the successful ingest of digital objects into the UHDL (or other access system depending on project specifications), preservation administrators prepare this content for transfer to UH Libraries’ digital preservation system by ensuring that it complies with SIP Specification (see Appendix B, Section 3.c: [“Ingest”](#) for additional information).

With born-digital items, steps are taken to insure the authenticity and integrity of the files, such as the use of write blockers or capture of disk images, during transfer of the files to the archives. Files are previewed before being transferred and validated within the archival setting. The information producer is asked to retain a copy of the files until they are ingested into the digital preservation system.

The Digital Preservation Working Group will determine workflows and revisions to policies for other materials that require preservation but have no existing policy, such as outsourced digitization or electronic serials produced by the University.

**Related Documentation:**

- [AIMS Donor Survey](#)
- [Special Collections’ Procedures for Accessioning Born Digital Content](#)
- Special Collections’ [Deed of Gift Form](#) (PDF), [Attachment A](#) (PDF), and [Attachment B](#) (PDF)
- [University Departmental Transfers](#)

### ***Note on confidential and sensitive data***

As defined in UH Campus IT's [Data Classification and Protection policy](#) (PDF), Level 1 data includes "confidential information" and "sensitive personal information." Confidential information includes social security numbers, educational records as defined by FERPA, health care information as defined by HIPAA, and customer information as defined by the Gramm-Leach-Bliley Act. Sensitive personal information includes an individual's first and last name or first initial and last name in combination with their social security number, driver's license or other government-issued ID number, and/or financial account numbers or login information.

Files in which it can reasonably be assumed confidential or sensitive data may appear, such as born-digital personal archives, are required to be scanned/searched using the best tools and techniques available at the time of processing. The following data will be redacted from preservation masters whenever possible prior to Archivematica ingest: SSNs, confidential educational or health care records, driver's license or other ID numbers (in combination with the person's name), and financial information.

## 3.b Common Services

### ***Operating system services***

The digital preservation system Archivematica will be run on the Ubuntu Operating System per the developers' recommendation. Ubuntu is a Debian-based Linux operating system that uses the open source development model. It is a mature product with a large installed base and an active development community. Ubuntu has a large and extensible suite of utilities, services, and applications that fully support the digital preservation system. Our Ubuntu server is run in a virtual machine from within the Libraries' highly scalable and highly available server virtualization environment.

### ***Network services***

Archivematica has full access to the Libraries' 1 Gbps switched Ethernet network using TCP/IP for communication. It will access common network services such as DNS, SMTP, etc. The digital preservation system will access the Libraries' storage area network using the iSCSI protocol and will access cloud storage locations using HTTPS over the Internet.

### ***Security services***

Shell level access to the operating system is limited to the Library Technology Services staff who are responsible for administering the server. Non-LTS staff access is granted on a case-by-case basis. Shell accounts are created and administered by LTS staff. Sudo is used to assign temporary privileges when performing administrative tasks in the shell.

All web interfaces that require a login are secured with 2048 bit SSL certificate. Accounts for the Archivematica web interface (dashboard) are limited to staff who need access to perform their roles in the preservation workflow. Archivematica allows two types of accounts, users and

administrators. Administrator accounts have full access to all the functions of the system. User accounts have access to all the functions of the system, except they cannot create other users and cannot modify the preservation planning settings. Dashboard accounts are created and administered by LTS staff. Access to, and administration of, the Archivematica storage services dashboard is limited to LTS staff and Digital Preservation Administrators.

All unnecessary operating system services are disabled. The built-in software firewall blocks all incoming network connections except those required for the administration and normal operations of the digital preservation system.

Security patches and updates are applied to the operating system each month or as critical patches are released. Key users will be notified when updates are scheduled for installation and if a server restart is required. Updates to the digital preservation system software are applied as released by the developer after thorough review by LTS staff and the DPWG.

System health, performance, security will be actively monitored by the Libraries' Xymon Monitor service which alerts LTS staff when events occur.

Preservation data will be backed up on a weekly basis to the Libraries' rdiff-backup server and retained for 30 days.

### 3.c Ingest

#### ***Preparation of digital assets for transfer***

At UH Libraries, digital assets that are candidates for preservation originate/reside in one of two departments: for digitized assets, MDS serves as a producer or steward of content; for born-digital assets, Special Collections plays a critical role through its accessioning of born-digital materials from donors or campus entities. Thus, to ensure a chain of custody for our digital assets, preservation responsibilities reside in both departments.

In preparing materials for transfer, both MDS and Special Collections create submission information packages (SIPs) that contain the materials necessary for the long-term preservation of digital assets. To ensure consistent quality across all submissions, the Digital Preservation Working Group has created SIP specifications aimed at preserving:

- The digital assets themselves,
- The minimal descriptive metadata attributed to these assets that is necessary to ensure appropriate levels of reliability, authenticity, and provenance,
- The structural metadata necessary to preserve a canonical record of original order and/or hierarchy (if applicable).

All metadata preserved in the preservation repository adheres to thoroughly-documented, widely-adopted standards that are executed in languages that are both human and machine readable for persistent access across time and changing technologies.

The Digital Preservation Working Group will determine workflows and revisions to policies for other materials that require preservation but have no existing policy, such as outsourced digitization or electronic serials produced by the University.

### ***University of Houston digital preservation system SIP specification***

This section outlines the minimum requirements necessary for all SIPs to be ingested.

#### Directory Structure & File Placement

A top-level folder for each digital object, complying with local naming conventions for top-level folders (either born-digital or digital conventions), will contain an “objects” directory, a “metadata” directory, and an optional “service” directory.

#### *Objects directory*

All preservation master files are placed in the objects directory. If there are modified masters or mid-level files, such as DVs for digitized video content, they are placed in the service directory.

For born-digital collections, the objects directory contains a file/folder structure that faithfully represents the original order/structure of the digital materials, using the original names given to folders by the creator(s). An EAD finding aid representing the files as they have been arranged and described for access, if available, will be placed in the submissionDocumentation folder. All spaces or other forbidden characters will be replaced by appropriate characters during the Archivematica transfer process.

#### *Metadata directory*

The completed metadata.csv is placed in the metadata directory. This file contains core descriptive metadata and identifiers, such as the ArchivesSpace URI of the object and the controlled vocabulary ARK for the collection title.

#### *Supplemental Documentation*

If there are donor agreements, transfer forms, copyright agreements, or any correspondence or other documentation relating to the transfer, a folder within the “metadata” directory named “submissionDocumentation” contains those files. “submissionDocumentation” is also where any metadata in other formats not currently supported by Archivematica’s ingest function are stored.

#### Core Metadata Record

The core metadata record for preservation objects will typically contain the following fields:

- dcterms.title
- dc.date
- uhlib.aSpaceUri
- dcterms.identifier
- dcterms.isPartOf
- uhlib.note
- partOfAIC

Of these elements, only `dcterms.title` is required. The rest of the elements are recommended.

For guidance on what information should be placed within these fields, refer to the [UH BCDAMS-MAP](#).

For information on how to construct the `metadata.csv` record for ingest, see [Metadata import](#) in the Archivematica user documentation.

### ***Archivematica ingest of SIPs and AIP/DIP generation***

During the ingest process, digital objects are packaged into SIPs and run through a range of microservices, potentially including normalization. In Archivematica, “normalizing is the process of converting digital objects to preservation and/or access formats.” Though this process results in new files, the original objects are always kept.

After the SIP is approved by the system, the package is run through additional microservices, which includes the processing of submission documentation, generation of the Archivematica METS file, indexing of metadata, and packaging of the AIP.

The Digital Preservation Working Group will determine workflows and revisions to policies for DIP content.

### ***Retrieving and updating AIPs***

In order to effectively track ingests, they must be logged in accordance with departmental documentation standards. A key piece of information that must be documented by all departments is the ARK that will be assigned to each preservation package during the digital access and preservation workflow. This will aid in AIP retrieval in the future. Currently, the only way to find and retrieve an AIP from Archivematica is through its search feature, which utilizes a limited index of digital asset metadata. Once an AIP is created, quality assured, and passed to archival storage, the package is not to be altered.

### **Related Documentation:**

- [Archivematica User Manual 1.6](#)

## 3.d Archival Storage

The storage locations used by the digital preservation system are volumes on the Libraries’ storage area network (SAN). The Libraries’ SAN is composed of multiple storage arrays that each have a large number of disk drives. The drives are managed by high performance redundant hard drive controllers in a RAID 6 configuration. This RAID level allows for the failure of up to 3 disks in each array before data is lost. The disk drives have predictive media-error detection and correction capabilities.

Collection files are prepared and temporarily stored on SAN storage volumes designated for that purpose. Users with accounts in preservation system dashboard can transfer collection files of various types into the system and create SIPs for ingestion into archival storage. Once successfully transferred and ingested into the archival storage, the collection files are deleted from the temporary and processing locations. SIPs and AIPs are stored in predefined locations in the preservation system storage that can be selected during processing.

Copies of files in archival storage are duplicated to offsite cloud storage location(s) periodically at the completion of the digital project workflow. Copies of archival storage files will be retrieved from the cloud storage locations in the event that local files are lost, corrupted, or backups don't contain the needed files.

Fixity checking of collections in archival storage is performed on a monthly schedule using Archivematica's fixity tool. A random sample of checksums will be checked each month, and a full fixity check of all checksums will be conducted during the 3-year audit.

Users with accounts in the preservation system dashboard can create DIPs from stored AIPs by browsing and downloading collections as needed.

### 3.e Data Management

AIPs placed in archival storage will not be changed or modified. Adding or changing objects in a collection is accomplished by ingesting new files, creating a new SIP, and relating it to the original AIP with the accession number in the package title.

Simple queries for filenames and metadata contents can be performed using the search function in the dashboard.

### 3.f Administration

Management of UH Libraries' digital preservation function will be a collaborative process, led by the Digital Preservation Working Group in conjunction with key library stakeholders. This section reviews the major services related to the system's implementation and functionality, provides the guidelines that inform these services, and identifies the agents who are responsible for implementing the services. It also addresses physical access and environmental control issues important to the long-term maintenance of electronic media.

**Chart 4: Administration of UH Libraries Digital Preservation System**

Service	Description	Guidelines	Responsible Agent
Establish Standards and Policies	Establish and maintain the UH Libraries Digital Preservation Policy and standards	As workflows and technology change, so too will the need for updating current standards as well as establishing new policies. The Digital Preservation Working Group will work closely with the UH Libraries' key stakeholders to monitor professional developments, craft changes to the Digital Preservation Policy, and coordinate the integration of these changes into existing workflows and programs.	DPWG
Manage System Configuration	Provide system specifications for the repository to continuously monitor the functionality of the entire repository and systematically control changes to the configuration	<p>Changes to the UH Libraries' digital preservation system are made through Archivemata's Dashboard Administration Tab. This interface controls content related to:</p> <ul style="list-style-type: none"> <li>● Processing Configuration</li> <li>● Failure Reporting</li> <li>● Transfer Source Locations</li> <li>● AIP Storage Locations</li> <li>● User Administration</li> </ul> <p>The Digital Preservation Working Group will determine default values for these content areas.</p> <p><b>Related Documentation:</b> Archivemata, <a href="#">Dashboard Administration Tab</a></p>	<p>DPWG</p> <p>Head of Library Technology Services</p>
Negotiate Submission Agreement	Solicit desirable archival information for the DPR and negotiate Submission Agreements with producers	<p>SIP data intended for ingest into Archivemata should comply with the specifications for digital objects and associated metadata, outlined in A.2 "<a href="#">Digital Assets</a>" and in B.3 "<a href="#">Functional Entities</a>" of the UH Libraries' Digital Preservation Policy.</p> <p>The Digital Preservation Working Group, in conjunction with the Digital</p>	Digital Projects Coordinator

		<p>Collections Management Committee, will determine the workflow for objects that fall outside this scope.</p>	
<p>Update Archival Information</p>	<p>Provide a mechanism for updating archive contents</p>	<p>Archivematica will preserve AIPs in the condition in which they were ingested into the system. Changes to files previously ingested will be submitted as new SIPs and associated with the previous version of the file using Archivematica's AIC function.</p>	<p>Digital Projects Coordinator</p>
<p>Repository Self-Audit</p>	<p>Verify that submissions meet the specifications of the Submission Agreement; will maintain a record of event-driven requests and periodically compare it to the contents of the archive to determine if all needed data is available</p>	<p>In the final year of each three-year cycle, the Head of Digital Research Services, the Head of Library Technology Systems, the Digitization Services Coordinator, and the Digital Projects Coordinator will conduct an audit of the UH Libraries' digital preservation system and its contents. During each review cycle, this group will establish digital preservation audit criteria, such as:</p> <ul style="list-style-type: none"> <li>● Content inventory</li> <li>● Core metadata</li> <li>● PREMIS event recording</li> <li>● Selection criteria</li> <li>● Emerging standards</li> </ul>	<p>DPWG</p> <p>Head of Digital Research Services</p> <p>Head of Library Technology Services</p>
<p>Provide Customer Support</p>	<p>Create, maintain and delete administrator accounts</p>	<p>Because digital objects are at great risk for degradation or loss, their exposure to human interaction should be minimized during the preservation process. Consequently, only key stakeholders with digital preservation responsibilities will have access to Archivematica. These positions include:</p> <ul style="list-style-type: none"> <li>● Head of Digital Research Services</li> <li>● Head of Metadata and Digitization Services</li> <li>● Head of Library Technology Services</li> <li>● Digital Projects Coordinator</li> </ul>	<p>Head of Library Technology Services</p>

### **Physical access and environmental control**

Ensuring a secure and stable environment for electronic media is essential to the long-term preservation of digital objects. Key units in UH Libraries have implemented security measures to prevent media from damage or theft. Servers in the Library Technology Services department are positioned in one physical location. UH Libraries restricts access to these areas through door locks and card-access devices. Those employees with direct reporting lines to Library Technology Services receive access to these respective areas.

In addition to secure physical locations, environmental conditions also play an important role in long-term digital preservation. Strict environmental controls are necessary to slow the rate of deterioration, since even electronic media is affected by the levels of temperature, relative humidity, light, and air pollution in which they are stored. Establishing baseline temperature and relative humidity controls and minimizing fluctuations in these areas slows chemical deterioration. Best practices recommend the following values for areas containing magnetic media for long-term digital preservation.

<b>Chart 5: Environmental Controls for Magnetic Media</b>			
<b>Temperature (Degrees F)</b>	<b>Allowable Range (+ or -)</b>	<b>Relative Humidity</b>	<b>Allowable Fluctuations (+ or -)</b>
65°	5°	30%	5%

### 3.g Preservation Planning

#### **Strategic priorities for digital preservation, 2015-2018**

Preparing for the demands of digital preservation requires UH Libraries to be strategic with its allocation of resources and talent. This section identifies the future actions of a digital preservation program, based on principles in the policy framework, and categorizes them as short-term (within the next three years) and long-term priorities. These principles focus on four key areas:

1. Repository robustness
2. Risk management
3. Planning and development
4. Collaboration

<b>Chart 6: Short-Term Priorities</b>	
<b>Category</b>	<b>Action</b>
Repository Robustness	<ul style="list-style-type: none"><li>• Normalize higher risk files (while retaining the original file) to preservation standards for those file formats where open specification or less risky file formats exist. As part of its implementation work, Digital Preservation Administrators will</li></ul>

	<p>jointly determine which files should be converted and to which particular file format to best provide characteristics such as functionality, longevity, and preservability. Archivemata will automate these preservation decisions.</p> <ul style="list-style-type: none"> <li>● Continually identify and implement essential preservation tools within the long-term digital repository, so that we can use them to reliably preserve our collections for future re-use.</li> <li>● Devise preservation plans for all major types of digital collection content held in the repository, so that we can invoke the necessary preservation tools in a timely manner.</li> <li>● Monitor file integrity, so that we may identify corrupt files and act accordingly to ensure only files with their integrity intact are delivered to users.</li> <li>● Utilize shared technical digital preservation services where appropriate such as representation information registries, so that we do not unnecessarily duplicate efforts.</li> <li>● Audit the repository against a recognized digital preservation repository audit methodology, such as ISO 16363:2012, so that we may independently validate our approach and measure our progress over time.</li> </ul>
<p>Risk Management</p>	<ul style="list-style-type: none"> <li>● Clearly define our technical requirements and collection policies for preservation throughout the information lifecycle, so that we can ensure preservation needs are known and can be addressed as and when relevant.</li> <li>● Implement rigorous quality assurance process for digitized content, so that we can identify content of inadequate quality before it enters the preservation workflow.</li> <li>● Implement rigorous quality assurance process for digitized content, so that we can identify content of inadequate quality before it enters the preservation workflow.</li> <li>● Implement tools and end-to-end workflows for digital content, so that we control the risks associated with receiving, managing, processing, and ingesting digital collection content.</li> <li>● Ingest valid legacy digital content into preservation storage as soon as possible, so that distributed and inconsistent storage and management practices are minimized and the risks associated with such practices addressed.</li> </ul>
<p>Planning and Development</p>	<ul style="list-style-type: none"> <li>● Document our relevant policies, procedures, standards, and systems development, so that they may be sustained, audited, and understood over time.</li> <li>● Plan and budget for long-term preservation of content at point of acquisition, so that financial sustainability is considered early in the life cycle.</li> <li>● Consider sustainability in all future system procurement exercises and content oriented partnerships, so that we enter new initiatives</li> </ul>

	<p>with a long-term vision and plan.</p> <ul style="list-style-type: none"> <li>• Ensure that all staff working with a responsibility for digital content understand the issues associated with preserving it, so that sustainability and preservation become an embedded consideration when developing and planning new systems and workflows.</li> </ul>
Collaboration	<ul style="list-style-type: none"> <li>• Seek out appropriate opportunities to collaborate with other institutions and organizations on digital preservation initiatives that meet our business needs, so that we may benefit from shared resources available to address shared challenges.</li> <li>• Ensure our collaboration with professional digital preservation membership organizations such as the National Digital Stewardship Alliance (NDSA) and the Digital Preservation Network (DPN) is in line with organizational requirements, so as to achieve the maximum return on investment in terms of time, effort, and financial commitment.</li> </ul>

<b>Chart 7: Long-Term Priorities</b>	
<b>Category</b>	<b>Action</b>
Repository Robustness	<ul style="list-style-type: none"> <li>• Test different technical strategies such as migration and emulation so that we can identify appropriate large scale approaches and tools to combat technological obsolescence.</li> </ul>
Risk Management	<ul style="list-style-type: none"> <li>• Integrate digital preservation risk management into our collection management and risk management strategies, so that digital risks are treated comparably with those facing analog content and regular preservation risk assessments are undertaken.</li> </ul>
Collaboration	<ul style="list-style-type: none"> <li>• Deliver successful contributions to collaborative projects already underway, including the NDSA, so that we meet existing commitments and maintain the Libraries' place at the cutting edge of international collaborative digital preservation.</li> <li>• Exchange knowledge and expertise across the wider international digital preservation and digital cultural heritage communities, for other institutions to learn from our work and provide opportunities to identify potential future partners with similar interests.</li> </ul>

***Content selection strategies for cloud-based storage***

As digital preservation activities increase at UH Libraries, the need to prioritize the type of content redundantly stored to the cloud may arise. We believe that everything preserved in Archivematica storage should be synced to DuraCloud. However, in the event that resources will not allow for everything to be transferred to DuraCloud, one potential model for prioritization

analyzes the level of risk certain kinds of digital and analog formats face and ranks them according to their current and future physical condition. UH Libraries' digital preservation stakeholders should review this model periodically, update it as resources and circumstances change, and be transparent when it moves to this model for cloud-based storage selection.

<b>Chart 8: Risk Levels for Content Selection</b>		
<b>Risk Level</b>	<b>Description</b>	<b>Type of Content</b>
Extreme Risk	No other copies of the digital or analog content are preserved.	<ul style="list-style-type: none"> <li>• Born-digital: unique holdings</li> <li>• Physical content: no analog resource available</li> <li>• Physical content: analog resource to fragile or cost-prohibitive to re-digitize</li> </ul>
Significant Risk	Analog content exists but additional digitization beyond an initial capture could result in damage to the item. Content could also be residing on formats that will likely become obsolete in the near future.	<ul style="list-style-type: none"> <li>• Magnetic tape and film content: resource in poor condition</li> <li>• Digitized content: obsolete file formats</li> </ul>
High Risk	Analog content exists but the material to be digitized is fragile or at a high preservation risk (e.g., magnetic tape). Re-digitization is not out of the question, though it is not ideal.	<ul style="list-style-type: none"> <li>• Born-digital: available elsewhere</li> <li>• Magnetic tape and film content: analog resource in fair condition</li> <li>• Physical content: analog resource in poor condition</li> <li>• Transparencies and photographic film: no photographic prints available</li> </ul>
Low Risk	Analog content exists and there are few limitations on the number of times an object can be re-digitized	<ul style="list-style-type: none"> <li>• Physical content: analog resource in stable condition</li> </ul>

### 3.h Access

The vast majority of digitized content is made available to consumers through the University of Houston Digital Library. The repository's contents are indexed by all major search engines, as well as by UH Libraries' discovery platform. Additionally, the UHDL interface allows consumers to search for content using simple and advanced search features. Browsing content is also available through the "Browse the Collection," "Related Collections," and "Related Items" features. Consumers are permitted to download certain static image files through the UHDL Digital Cart Service.

Currently, consumers determine the availability of born-digital content stored in the UH Libraries digital preservation system through the online finding aid. For collections in which the digital materials have already been processed and described in the finding aid, access files will be available for viewing on the Reading Room computer. Consumers may also contact the collection curator to request and receive information products that are served up in the reading room on demand.

## 4. References

Archivematica 1.6 Documentation, 2017

<http://www.archivematica.org/en/docs/archivematica-1.6/contents/>

British Library, Digital Preservation Strategy, 2013

[http://www.bl.uk/aboutus/stratpolprog/collectioncare/digitalpreservation/strategy/BL\\_DigitalPreservationStrategy\\_2013-16-external.pdf](http://www.bl.uk/aboutus/stratpolprog/collectioncare/digitalpreservation/strategy/BL_DigitalPreservationStrategy_2013-16-external.pdf)

Lyrasis, Environmental Specifications for the Storage of Library & Archival Materials, 2009

<http://www.lyrasis.org/LYRASIS%20Digital/Documents/Preservation%20PDFs/environspec.pdf>

Gail McMillan and Rachel Howard, "Chapter 5: Content Selection, Preparation, and Management," A Guide for Distributed Digital Preservation, MetaArchive Cooperative Publications, eds. Katherine Skinner and Matt Schultz, 2010

[https://educopia.org/sites/educopia.org/files/publications/A\\_Guide\\_to\\_Distributed\\_Digital\\_Preservation\\_0.pdf](https://educopia.org/sites/educopia.org/files/publications/A_Guide_to_Distributed_Digital_Preservation_0.pdf)

# Appendix C: Digital Preservation Glossary

Access: The ability, permission/right, and means to locate, display, obtain, determine availability of or make use of a digital asset, or information about that resource.

Archival Information Collection (AIC): OAIS reference model defines AICs as an Archival Information Package whose Content Information is an aggregation of other Archival Information Packages.

Archival Information Package (AIP): AIPs consist of Content Information and the associated Preservation Description Information (PDI), which is preserved within the digital preservation repository.

Authentic copies: A duplicate of a digital asset that is what it purports to be and that is free from tampering or corruption.

Authenticity: A quality of a digital asset to be judged trustworthy and genuine based on internal and external evidence.

Digital Preservation: The whole of the activities and processes involved in the physical and intellectual protection and technical stabilization of digital asset through time in order to reproduce authentic copies of these resources.

Digital Resource/Asset: Encoding of intellectual or cultural context in digital form.

Dissemination Information Package (DIP): DIPs are derived from one or more AIPs and received by the consumer in response to a request to the digital preservation repository.

Open Archival Information System (OAIS) Reference Model: A conceptual framework for an archival system dedicated to preserving and maintaining access to digital information over the long term. Within the OAIS model, three types of information packages are identified: the Submission Information Package (SIP), the Archival Information Package (AIP), and the Dissemination Information Package (DIP).

Provenance: The source and ownership history of an asset.

Submission Information Package (SIP): SIPs are delivered by the producer to the digital preservation repository for use in the construction of one or more AIPs.

Trustworthy: Being able to provide reliable, long-term access to managed digital resources to a designated community, now and in the future. Signs of trustworthiness include organizational,

administrative, technical, and financial viability along with system security. This term is rooted in the concept of the Trusted Digital Repository.

“Trusted Digital Repositories: Attributes and Responsibilities”

<http://www.oclc.org/content/dam/research/activities/trustedrep/repositories.pdf>