

Yale University Library Policy for the Digital Preservation

Yale University Library (YUL) Mission:

The Yale University Library, as one of the world's leading research libraries, collects, organizes, preserves, and provides access to and service for a rich and unique record of human thought and creativity. It fosters intellectual growth and supports the teaching and research missions of Yale University and scholarly communities worldwide.

Digital Preservation:

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

Digital Preservation Principle Statement:

Yale University Library Digital Preservation Policy supports the preservation of digital resources that are within the Library's collections.

These digital resources are subject to the same criteria for selection and preservation as other resources in the Yale Libraries. These decisions are made by selectors, curators, and bibliographers as experts on the value of the content, in consultation with the relevant information technology and preservation experts. Digital preservation decisions are made on the basis of this Policy, the Library's Strategic Plan, the digital resources' enduring value and the feasibility of the digital resources' preservation. When possible, decisions about the need for preservation are made at the time of creation, acquisition, or licensing of digital resources.

Selectors, curators, and bibliographers in consultation with technical experts must specify the preservation requirements for the digital resource. Preservation responsibility is retained by YUL whether the digital resource is preserved at YUL or entrusted to an outside agency. Preservation of digital resources may include any actions necessary to preserve continued access to the digital material, ensure its authenticity, mitigate and/or reverse the effects of hardware and software obsolescence and media decay.

This Policy recognizes that the maintenance and the reliable long-term access to Yale's digital resources are supported by a preservation planning function. Research (monitoring) about technology that supports a repository and the requirements of the designated community it serves is a core activity to preservation planning, as well as outreach and education regarding policies, procedures and best practices for digital resources.

Frequency of Policy Review

This policy will be reviewed annually to assure timely updates to reflect the maturing of the technology by the relevant body.

Identification of Content, File Format, Source and Collecting Levels of Digital Resources to be Preserved:

For digital resources the decision to preserve, as noted above, is based upon their enduring value and the feasibility of preservation and not necessarily upon a digital resource's content, format type, the source of the resource or the collecting level. This policy recognizes that preservation strategies and actions vary by these attributes and characteristics. Below are four tables that identify examples of content types, formats, sources of digital resources and collecting level that are likely to be represented in a preservation repository at YUL or entrusted to a third party to preserve upon Yale's behalf.

Digital Content Type Examples

Content Type	Example
Text Resources	Theses, dissertations
Resources generated using office applications	Spreadsheets, Power Point presentations
Electronic records	emails, financial and administrative documents
Simple or complex audio, video and image files	Speech recordings, conference videos, high-resolution images of works of art
Datasets	Census data, epidemiological study data
Web resources	Yale University Frontdoor

Digital Format Examples:

The format landscape is ever changing as new formats emerge overtime. Examples of digital formats include:

Content Type	Format Example
Text	ASCII (ANSIX3.4 ECMA-6, ISO646), UTF-8 Unicode
Audio (voice, music)	AIFF, Wave, MP2
Video	MPEG, MPEG-2, AVI
Image	Gif (Gif87a, Gif89a), JPEG, JPEG2000, TIFF (TIFF4.0, TIFF5.0, TIFF6.0)

Sources of Digital Content include:

- o Born digital resources and digital surrogates created by YUL.
- o Born digital resources and digital surrogates originating external to YUL; for example other university departments, for which YUL has a mandate to preserve.
- o Digital resources acquired by YUL through purchase or donation.
- o Digital resources licensed by YUL with perpetual access and archival rights.

Collecting Levels:

Collecting levels for digital resources vary considerably. Examples include resources hosted at Yale, resources hosted at Yale and elsewhere, i.e., mirrored resources, and resources hosted elsewhere that YUL makes available through links to the hosting location. Not all levels of digital resource collecting include an automatic commitment for preservation, although any of these resources may be selected for preservation.

Regional, national, international and consortial activities:

YUL acquires digital resources created commercially and by individuals and/or groups outside of YUL. The preservation of such digital resources will be considered the responsibility of another institution or organization if metadata and functionality are preserved at least as well as they would be by YUL. When considering preservation of these digital resources by another entity, YUL must regularly review their policies and practices to ensure our standards are maintained.

Life cycle:

Throughout the life cycle of digital resources a series of inter-related strategic and procedural decisions and the work of a number of different stakeholders contribute to help ensure that digital resources survive through time and changing technologies. The information life cycle is a framework for understanding the cyclical sequence of activities that all digital resources undergo during their existence. Digital resources may be conceived; born or adopted; utilized, shaped, and molded; protected from harm. Eventually digital resources may reach the end of their active lives and be disposed of in some manner, or be re-born as reformatted, transformed resources.

Although this Policy primarily focuses upon the preservation stage of the digital life cycle, the prospects for and the costs involved in preserving digital resources through time and technological change rest heavily upon decisions taken about those resources at different stages of their life cycle. By adhering to a pro-active concept of preservation management, we seek

minimal loss in digital information content, functionality, and accessibility and seek to ensure that digital resources retain qualities of integrity, authenticity and reliability over time.

Storage:

Libraries, museums or academic departments at Yale that provide preservation or archival services for their designated communities will by necessity purchase large quantities of storage for digital resources over time. The support of large scale storage is complicated and requires major investments in technology and staff to efficiently maintain and operate a storage management system that delivers basic services such as access, backup and disaster recovery. This Policy recognizes that there may be multiple archives or repositories at Yale that have independent storage systems. The sheer volume of storage needed to archive Yale's digital resources makes the economic and operational management of storage an important issue for the University.

Yale is best served when distributed and disparate systems conform to standards and best practices that make communication between these storage systems possible. The ability to integrate or interoperate within and between storage systems is likely to make backup, disaster recovery and hardware migration services less risky across all storage systems and more economical for the University. For example, common import and export services will enable one storage system to serve as a backup for another thereby reducing Yale's total investment in redundancy for these systems. In addition, to satisfy the custodial requirements of depositors for trust and security, the management of storage will require specialized technology, software and management skills. To ensure that Yale makes the right investments in storage for digital preservation this Policy recommends that the acquisition and management of storage be guided by technical considerations that are discussed in the best practice section of this Policy (under development). This Policy also maintains that storage technologies must be seen as chronically obsolescent and subject to continuous hardware migrations over time.

Authenticity:

YUL strives to ensure the authenticity of digital resources; the mutable nature of digital resources opens the possibility for unauthorized and undetectable changes. Confidence in the authenticity of digital resources over time is particularly crucial owing to the ease with which alterations can be made. From the moment that digital resources are created or acquired, YUL undertakes protective procedures to prevent, discover, and correct loss or corruption of digital resources due to either inadvertent or malicious intent. In addition, supporting evidence, ideally in the form of metadata, must be provided to enable users to evaluate the authenticity of all preserved digital resources.

Metadata:

Metadata is fundamental to preserving Yale University Library's digital resources. Preservation metadata includes a number of different types of metadata: administrative (used in managing information resources including rights and permissions), technical (describing hardware and software needed to maintain an information object) and structural (identifying the relationships between objects such as *part of*, *dependent upon* that form intellectual entities. Particular attention is paid to the documentation of digital provenance (metadata documenting the history of the object and any actions taken to maintain and provide access), and of relationships among different objects within preservation repositories (vs. relationships between resources, i.e., structural metadata).

The preservation process must be able to understand, take in, and maintain metadata submitted with the digital resource while creating its own metadata to manage the preservation of that resource. For policies, procedures and best practices related to the creation and handling of metadata at YUL see the IAC Metadata Committee's website at: <http://www.library.yale.edu/cataloging/metadata/IACmetadata.html>. For specific policies,

procedures or guidelines regarding the creation and maintenance of preservation metadata see the IAC Preservation Metadata Committee's website (under development).

Access:

In preserving the accessibility of digital resources, the Library will:

- o Maintain information regarding rights and permissions governing access.
- o Maintain the means of accessing an acceptable presentation of the digital resource; and
- o Maintain the ability to locate the digital resource reliably.

Intellectual Property:

The preservation of a digital resource will include complying with the Intellectual Property rights and/or other legal rights related to copying, storage, modification and use of the specific resource. (See Yale Office of Cooperative Research, http://www.yale.edu/ocr/indust_policies/).

Resource Management:

Enduring preservation of digital resources requires substantial and ongoing resource management over time. Digital preservation is dynamic; responses to technological obsolescence or media decay must be taken more quickly and the life expectancy of a preservation treatment is shorter because the technologies utilized are evolutionary. Consequently, preservation strategies must be periodically monitored and reassessed as the technological environment that supports standards, protocols, and formats, etc. evolves.

While the overall resource commitments to digital preservation are understood to be substantial, the exact costs (staff and dollar) of preserving digital resources over time are now difficult to identify and define. Resources are needed throughout the life-cycle of these materials for such activities as: ([Digital Cost Center URL](#))

- o Creation, acquisition, evaluation and selection relating to preservation; including, but not limited to: determining the preservation strategy and negotiating for right to preserve resources as well as the rights management policies.
- o The technical infrastructure, i.e., hardware, operating systems, software, network, physical facility, etc. and one-time resource commitments for the purchase of these components as well as recurring commitments for licensing fees, maintenance charges, facilities and supplies.
- o Administration, i.e., resources for administering the preservation technical environment, overall digital preservation program, related outreach throughout the University, following relevant developments and laws pertaining to digital preservation, conducting research, financial planning and outsourcing.
- o Data preparation and validation, i.e., resources related to implementing preservation strategies and obtaining any necessary documentation, available metadata, machine validation and authentication of the digital resource.
- o Data management, i.e. resources, through time, associated with managing digital resources such as format migration, digital resource deletion, re-validation, re-authentication, preservation metadata creation and maintenance.

More centralization of digital preservation will help reduce the overall resource commitment by integrating activities and exploiting economies of scale. All YUL administrators, creators and collectors of digital resources are stakeholders in digital preservation and remain aware of the resource implications of their activities and decisions.

Best Practices:

This policy is supported by Yale University Library's digital preservation best practices. (In development FY07)

Glossary:

This glossary offers non-technical definitions of terms used within this policy.

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

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

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

Content: The material, information and intellectual substance of a digital resource.

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

Digital resources: Encoding of intellectual context in digital form.

Enduring value: The continuing usefulness or significance of digital resources, based on the administrative, legal, fiscal, evidential, or historical information they contain and function they serve, justifying their on-going preservation. The phrase “enduring value” emphasizes the perceived value of the digital resources when they are appraised, recognizing that a future selector may reappraise the records and dispose of them.

File Format: The organization (fixed, byte-serialized encoding) of digital information according to a preset specification.

Intellectual Property: Intellectual property is intangible property that is created by the mind. Like tangible real or personal property, the law recognizes the right to own and to control intellectual property. There are four well recognized types of intellectual property rights: copyrights, trademarks, patents, and trade secrets. These forms of intellectual property differ significantly in the rights they confer, how they are obtained, and how they are maintained.

Life cycle is the framework for understanding the cyclical sequence of activities that all digital resources undergo during their existence.

Maintenance: (*of digital resources*) The action of keeping the components of digital resources in working order or in repair. This includes loading digital resources into storage, managing the storage hierarchy, refreshing the media on which digital resources are stored, performing routine and special error checking, providing disaster recovery capabilities, etc. Maintenance may be differentiated from the broader term Preservation because Maintenance does not include the metadata management, preservation planning, and access controls necessary to ensure intellectual protection and to reproduce authentic copies of the digital resources over time.

Metadata: Information that describes significant aspects of a resource. Preservation metadata are required to describe, manage and preserve digital resources over time and will assist in ensuring essential contextual, historical, and technical information that are preserved along with the digital resource.

Perpetual access refers to permanent use of publishers' retrospective backfiles for subscribed years to specific publications with content in the same format and access method with which the publisher provides current content. (See CDC, 16 Dec 1999, Expectations of Yale University regarding Creating Archives of and Perpetual Access to Electronic Resources, <http://www.library.yale.edu/ecollections/yalearchiving.pdf>).

Preservation is the whole of the activities and processes involved in the physical and intellectual protection of administrative, legal, fiscal, evidential, historical information and cultural materials. Preservation encompasses a host of policies, procedures, and processes that together sustain access or prevent further deterioration to the materials we choose to save.

Preservation repository: Technical infrastructure, policies, procedures and corresponding management services that provide for the storage, maintenance and preservation of data or information for long-term use and retrieval.

Provenance: The source and ownership history of a digital resource.

Stakeholder: A person or group with an interest, involvement or investment in the digital resource.

SOURCES:

This policy has been informed by the following sources:

Life-Cycle Management of Digital Data. Preservation Reformatting Division. Library of Congress.
<http://www.loc.gov/preserv/prd/presdig/preslifecycle.html>

Neil Beagrie and Daniel Greenstein, A Strategic Policy Framework for Creating and Preserving Digital Collections. Version 5.0 (Arts and Humanities Data Service Executive, 1998, updated July 2001) <http://ahds.ac.uk/strategic.pdf>

Shenton, Helen. Life Cycle Collection Management.
<http://liber.library.uu.nl/publish/articles/000033/article.pdf>

Digital Preservation Coalition Handbook. <http://www.dpconline.org/graphics/intro/definitions.html>

Columbia University Libraries. Policy for Preservation of Digital Resources. July 2000
<http://www.columbia.edu/cu/lweb/services/preservation/dlpolicy.html>

John Garret & Donald Waters, Preserving Digital Information, May 1996
<http://www.rlg.org/legacy/ftpd/pub/archtf/final-report.pdf>

National Library of Australia, Digital Preservation Policy, Feb 2002
<http://www.nla.gov.au/policy/digpres.html>

Yale Office of Cooperative Research, Yale Intellectual Property Policies
http://www.yale.edu/ocr/indust_policies/

PREMIS, Preservation Metadata Implementation Strategies
<http://www.oclc.org/research/projects/pmwg/>

InterParesThe Long-term Preservation of Authentic Electronic Records: Findings of the InterPARES Project <http://www.interpares.org/book/index.cfm>

The International Research on Permanent Authentic Records in Electronic Systems (InterPARES) 2 Project: Experiential, Interactive, Dynamic Records
http://www.interpares.org/ip2/ip2_index.cfm

Original policy established: 11/05
Updates: 2/07