

Dartmouth College Library

Digital Preservation Policy

Executive Summary

This document was created to describe the need and strategies for preserving Dartmouth College Library's digital resources. Rapid growth in both the number of digital resources and the proportion of the Library's budget used to obtain them necessitates that proactive steps be taken to preserve these materials. These digital preservation activities ensure that faculty, staff, students, and other users will have ongoing access to the Library's expanding digital collections.

This policy provides a broad set of guidelines for digital preservation, from which procedures can be developed with confidence that they will meet accepted standards, make effective use of resources, and support the mission and goals of the Library. Objectives of the policy are to:

- Describe the challenges associated with digital preservation.
- Explain why a digital preservation policy is necessary.
- Outline principles on which digital preservation actions will be based.
- Define the scope of digital preservation activities, including sources and types of digital content that will be preserved.
- Describe specific preservation strategies that will be performed to ensure the long-term preservation of digital materials. These strategies include life cycle management of resources owned by the College and the negotiation of third-party preservation agreements for licensed resources.
- Identify stakeholders responsible for components of the digital preservation strategies.
- Define a schedule for regular policy review.
- Define terms, identify standards, and list resources that will inform digital preservation activities.

In order to fully implement this policy, certain infrastructure requirements must be met. These include the addition of a full-time Digital Preservation Librarian to the staffing model, and the creation of a long-term digital preservation repository system.

Introduction

Library Mission and the Challenge of Digital Preservation

The Dartmouth College Library fosters intellectual growth and advances the mission of Dartmouth College and affiliated communities by supporting excellence and innovation in education and research, managing and delivering information, and partnering to develop and disseminate new scholarship.

Preservation Services is responsible for preserving the entire Dartmouth College Library collection to make it accessible for current and future students, faculty, and scholars. Continued long-term access to the Library's increasing digital collections is an essential component of the work Preservation Services performs to fulfill the Library's mission.

Dartmouth College Library has created digital content since the 1980s. With the start of the Digital Publishing Program in 2001 and the development of a digital production unit in 2008, digital production is now a daily activity. The Library also purchases and licenses a very large and growing number of digital resources. Due to the fragile nature of digital objects along with continually evolving hardware, software, standards, and file formats, these materials are at a much higher preservation risk than traditional analog materials.

Digital preservation is defined as “the series of management policies and activities necessary to ensure the enduring usability, authenticity, discoverability and accessibility of content over the very long term.”¹ Digital preservation differs from analog preservation in several ways. The primary difference is that digital preservation requires active management. While many analog materials, such as books, can survive for years when simply stored in a climate-controlled environment, digital materials that are left alone for long periods of time are much more likely to degrade beyond repair, and this degradation is generally not discovered until there is an attempt to use the item.

Additionally, the preservation needs of analog materials, such as books, journals, film, and tape, are well-understood and have not greatly changed over time. Digital preservation, however, is a new and developing field with standards that are still being created. New tools and technologies will require that digital preservation activities be responsive and adaptable.

Finally, the expertise to treat analog materials generally exists within one department; for the majority of the Library's physical collections, that department is Preservation Services. The expertise and actions required to preserve digital content exists across multiple departments, including Preservation Services, Digital Library Technologies Group, College Computing, Cataloging and Metadata, and others. A robust digital preservation infrastructure will inherently operate within a collaborative and communicative workspace.

¹ (JISC, Digital Preservation Coalition, Digital Archives Department of the University of London Computer Centre, Portico, 2009)

Mandate

A well-defined digital preservation policy is essential for the Library to carry out its mission of supporting excellence in research, delivering information, and disseminating new scholarship. Without a policy that defines its scope, strategies, challenges, and responsible parties, digital preservation will continue to be an ambition rather than a robust program. Reasons for developing a digital preservation policy include:

- Supports Recommendation 8 of the 2007 Dartmouth College Library Self Study, to “Develop a digitization infrastructure”.²
- Supports the 2010 *Dartmouth Digital Library: Program, Priorities, and Policies* report, which charges Preservation Services with “formulating long-term curation guidelines for the content built by the Digital Program”.³
- Supports the policies and procedures of Records Management⁴ and College Archives for providing access to digital records over their lifetime.
- Supports ongoing Library goals to develop a long-term repository for digital collections.⁵
- Supports ambitions of the Dartmouth Digital Information (D2I) committee to “plan for secure, long-term, preservation-aware storage of faculty scholarly output”.⁶
- Supports consortial agreements that carry a preservation responsibility across institutions.

This digital preservation policy is created in harmony with policies of the Digital Projects and Infrastructure Group (DPIG) and Collection Management and Planning Group (CMPG). The Head of Preservation Services is ultimately responsible for the implementation of digital preservation policies and procedures in the Library.

² (Dartmouth College Library, 2007)

³ (Digital Projects and Infrastructure Group, 2010)

⁴ (Dartmouth College Library, 2008)

⁵ (Dartmouth College Library, 2009)

⁶ (Dartmouth Digital Information, 2009)

Principles

The following principles will guide digital preservation actions:

Access: Digital preservation activities are performed with the understanding that long-term access is the primary goal. Access to digital collections will be supported to the best of our ability given available technology and resources, however perpetual access to digital materials cannot be guaranteed.

Authenticity: Digital objects will be created with supporting metadata to establish authenticity and provenance. Digital objects will be managed to ensure that they are unaltered and the original data is preserved.

Collaboration: Dartmouth College Library will investigate and participate in collaborative agreements whenever they are a good use of Library resources.

College and Library missions: This policy and actions taken to implement the policy exist in support of stated Dartmouth College and Library missions. The digital preservation policy will be annually reviewed against College and Library missions and goals to ensure that it continues to support the core work of the institution.

Intellectual Property: Dartmouth College Library is committed to providing access to digital materials while respecting and upholding the intellectual property rights of authors and obtaining prior consent when the creator's identity is known. Rights management actions will be documented and rights information will be preserved with digital content.

Standards and Best Practices: Dartmouth College Library will observe current standards and best practices related to the creation, maintenance, storage, and delivery of digital objects and metadata, as determined by international, national, consortial, and local institutions and governing bodies.⁷

Sustainability: Digital preservation activities will be planned and implemented in ways that best manage current college resources and can be sustained into the future. Future access to digital resources cannot be assured without institutional commitment to necessary resources.

Training: Dartmouth College Library will commit to on-going training and development of staff in areas related to digital preservation, as well as outreach to inform faculty, students, and staff of the best practices for creating and maintaining digital objects.

Technology: Dartmouth College Library will fulfill digital preservation objectives by developing and maintaining necessary hardware, software, expertise, and protocols to ensure long term access.

⁷ See Appendix 1 for detailed information on standards followed.

Scope

Content Sources

The Library selects, creates, and collects different types of digital resources. Digital resources collected by the Library fall into these general categories:

1. Subscription-based resources:
Digital files such as electronic journals, databases, digital collections, and e-books, to which the Library pays for access, but does not own outright.
2. Escrow files:
Archival backup files of electronic resources, which may be purchased by the Library or submitted by the vendor in fulfillment of contractual obligations. These may be housed at the Library or remotely. While they contain the complete contents of the digital resource these typically do not have the same “look and feel” as the original content. They may or may not be retained long-term depending on collection development policy decisions.
3. Dartmouth College-owned digital resources:
 - a. Analog objects owned by the Library that are selected for digital conversion either by Dartmouth College Library or a vendor.
 - b. Born-digital objects and publications created by Dartmouth College Library.
 - c. Electronic publications created by Dartmouth College faculty with the assistance of Library staff and hosted by the Library.
 - d. Digital files that are produced in the course of creating a physical exhibit and digital versions of selected Library exhibits.
 - e. Digital resources collected by Dartmouth College Library that are unlikely to exist elsewhere so should be preserved.
4. College records:
Records created by departments within the College in the course of conducting business. They will have varying retention schedules and should be considered confidential except to their creators or college administration.

Dartmouth College Library is committed to the preservation of all of these digital resources throughout their life cycle and will develop the technical infrastructure to support the creation, maintenance, and access of digital materials for the long term. It is also committed to supporting staff in developing the expertise to perform the activities.

Content Types

Each of the above content sources may present content in one or many of the following types, which may require different preservation strategies due to their varying attributes.

- Textual materials (ebooks, articles, etc.; ASCII, UTF-8, Unicode)
- Images (scanned books or photographs, digital photographs, digital art; TIFF, JPEG, GIF, JPEG2000)
- Audio/video materials (videos produced on campus, recorded sound oral histories, etc.; MPEG, AVI, MOV, AAC, WAV)
- Numerical data/datasets (research data; XML, XLS, proprietary database formats)

The library will likely acquire materials in additional formats in the future, and preservation strategies will be developed to accommodate new formats as needed.

Digital Preservation Strategies

Preservation Actions

The specific preservation actions used for Dartmouth College Library's digital resources will depend largely on the source and type of content, as well as existing technology, expertise, and ongoing support. Preservation actions based on current resources can be broken down as follows:

Subscription-based resources:

As these resources are not owned or directly controlled by the Library, they cannot be managed by Library staff. Instead, subscription-based digital resources are primarily managed by agreement with the publisher or vendor to use third-party preservation services (such as Portico and LOCKSS). The Library will negotiate such preservation agreements when developing subscription and license contracts with publishers and vendors.

The Library will also continue its participation in Portico, LOCKSS, and HathiTrust, in support of third-party archiving arrangements of resources not owned by the Library. The value of participation in these and other such services will be regularly assessed.

Resources created by or for, and owned by the College:

These resources will be comprehensively managed using the life cycle model outlined below. Expectation is that all Library-owned resource content and associated metadata will be developed according to current standards and best practices, and stored in a long-term repository within the Library infrastructure or in a consortium-based repository system (such as HathiTrust).

Life Cycle Management

Digital objects will be managed using the life cycle model⁸, which is a framework describing the stages that digital resources go through during their existence. The preservation of digital objects requires planning and action at every stage of an object's lifecycle, including each of the following areas:

Creation – As digital content is created, whether this is done by the Library or an external vendor, preservation actions should include creating and/or capturing administrative, descriptive, structural and technical metadata about the objects, as well as imposing a well-defined storage system. Content will be created following current standards and best practices for capture and formatting.

Selection – Selection for digital preservation will be done in coordination with current use, existing Library collection development policies, and collaborative agreements, while addressing specific format needs and budgetary limitations. All preservation actions will be taken under the assumption that materials selected for the library collections are intended for permanent retention unless explicitly stated otherwise.

Ingest – Ingest of materials into the collections will strictly follow local guidelines for ingest procedures. These guidelines will include delivery of content to the responsible department/personnel, verification of file types, validation of file content, normalization of files as needed, creation or enhancement of metadata according to standards set forth in local metadata policies, and transfer of data and metadata to an approved long-term storage system.

Metadata Creation – All digital resources created by the Dartmouth College Library will adhere to the Library's pending Metadata policy. Essential preservation metadata includes:

- Administrative
- Technical
- Structural
- Provenance
- Rights

Storage – Digital resources must be stored in a manner that is consistent with accepted best practices in the digital preservation community. This will include both technical infrastructure (hardware, software, network access, data backup, facilities, maintenance, etc.) and ongoing preservation management activities. Best practice in digital preservation requires duplicating digital objects in both local systems and geographically removed systems. Dartmouth College Library will pursue this by

⁸ (Digital Curation Centre, 2010)

working with College Computing to host redundant local storage. Library staff will also explore other methods of storing data off site, such as in a private LOCKSS network, the HathiTrust, the Internet Archive, or another collaborative group.

Preservation Management – A series of actions that will need to be performed on digital resources prior to and during long-term storage, at varying levels depending on the source and type of resource. Detailed procedures and workflows for preservation actions will be created and maintained. Possible preservation actions include, but are not limited to:

- Content and metadata validation
- Preservation audits – Preserved content will undergo periodic audits to ensure that activities are meeting stated commitments, that risks are reduced, and to verify authenticity and accessibility of content.
- Ongoing file format review
- Migration – conversion of data to new file formats and/or migration to new storage media as needed.
- Definition and monitoring of backup procedures.
- Maintenance of technical components such as hardware and software used for storage and access.

Access and Use - Digital objects and collections will be reviewed and managed to ensure that files are accessible into the future. Digital objects will be discoverable: created in a way that they may be easily found by all stakeholders.

Transformation – Digital resources may require periodic modification. Possible reasons for modification include: to support new developments in scholarly research capability, to function optimally in new delivery systems, and to prevent format, hardware, or software obsolescence. Types of modifications that may be performed include creating new content or metadata, adding content or metadata, migrating content to a new format, or creating a subset of content or metadata.

De-selection – Digital objects will be reviewed and disposed of as needed, based on collection development policies.

Stakeholders

Stakeholders in digital preservation include Library staff, users of Library collections (both at Dartmouth and elsewhere), faculty and other College staff who create digital content housed by the Library. Explicit responsibilities of stakeholders in carrying out preservation strategies include:

Acquisitions Services – Manages the purchasing and licensing of electronic resources.

Cataloging and Metadata Services – Manages the creation of metadata to ensure compliance with standards, best practices, and existing metadata policies.

Collections Management and Planning Group – Manages the collection development review and de-selection of digital resources as needed. Ensures ongoing harmony of digital collections with print collections and the Library's collection development policies.

College Archives – Selects and manages College records to be preserved.

Digital Library Technologies Group – In coordination with other Library departments and Computing Services, manages the technical infrastructure needed to create, ingest, store, transform, and provide access to digital resources. Creates, installs, and maintains software as needed and provides support for staff using these tools.

Digital Projects and Infrastructure Group – Manages the creation of digital content within the Library. Ensures that standards and best practices are followed for the creation of digital content, including the capture of preservation metadata.

Digital Resources Program – Manages the Digital Publishing Program and the licensing of subscription-based digital content. Ensures that sufficient third-party preservation agreements are met whenever possible.

Preservation Services – Oversees and manages the Library's digital preservation strategies, with particular emphasis on selection, ingest, storage, preservation management, transformation, and coordination with third-party preservation services. Ensures general compliance with standards and best practices. Coordinates activities across departments and with external vendors.

Records Management – Manages College records, including ingest of records into the records management system and subsequent transfer to College Archives or other storage as needed.

Web Steering Committee – Manages accessibility and user interface design to ensure usability and discoverability of digital resources.

Infrastructure Recommendations

In order to fully comply with this policy and the standards listed below, certain infrastructure components need to be implemented. Given current staffing levels, we do not have the resources to carry out all of the preservation strategies involved in life cycle management. The lack of a digital preservation repository is also a major impediment to compliance with basic standards and accepted practices. Therefore, we recommend the following be fulfilled before an attempt is made to completely implement this policy.

Digital Preservation Librarian

Given the need for a digital preservation repository and long-term preservation management of digital collections (articulated in the 2011 DPIG Program Plan⁹ and in FY12 Library Goal 2¹⁰), there is an immediate and pressing need for a full-time, professional staff position dedicated to the development of a digital preservation program for the Library. Current staff time allocated to digital preservation is less than 0.5 FTE, including both the Head of Preservation Services and the Preservation Specialist. This level of staffing will not be sufficient to support digital preservation as outlined in this policy. We recommend that a Digital Preservation Librarian be hired to take responsibility for this work.

Digital Preservation Repository

OAIS-compliant long-term storage that supports the life cycle model outlined above will require a robust digital preservation repository system. This system will provide the ability to perform ingest, metadata creation, archival storage, preservation management, and long-term access functions on digital collections. We recommend that work on developing this repository system begin as soon as possible, preferably after the appointment of a digital preservation librarian who will direct the process. A detailed outline of the desired digital repository can be found in the 2011 DPIG Program Plan appendices.

Policy Review

This policy and the actions that flow from it will be evaluated regularly to ensure that implemented strategies continue to support the Library's mission and policies, use resources in a cost-effective manner, and adapt appropriately to address evolving technologies. This evaluation will be completed at least once every three years.

⁹ Not yet published.

¹⁰ (Dartmouth College Library, 2011)

Preservation Services

9/27/2011

Dartmouth_digital_preservation_policy.pdf

Appendices

Appendix 1 - Standards and Best Practices

Dartmouth College Library will observe national and international standards and best practices for the creation and management of digital objects, along with the associated metadata needed to maintain resources throughout their lifecycle. Open source formats will be preferred.

Relevant standards include:

- Open Archival Information System Reference Model (OAIS)¹¹
- PREMIS Data Dictionary for Preservation Metadata¹²
- Trustworthy Repositories Audit & Certification (TRAC): Criteria and Checklist¹³
- Dartmouth College Library Guidelines:
 - Digital Policies and Procedures: File-Naming Conventions, Version 1.0
<https://libwiki.dartmouth.edu/twiki/bin/view/Libopen/DLGprocedure01>
 - Proposed Rights Policies for the Publication of ad Access to Digital Works Through Dartmouth Digital Library Programs: Digitized Collections and Digital Publishing (provide URL when published)
 - Records Management: Record Production and Maintenance FAQ.
<http://www.dartmouth.edu/~library/recmgmt/production.html>
- Dartmouth College Computing Guidelines:
 - Effective Data Management, Richard Brittain, John Wallace, Jaime Combariza, Research Computing, Dartmouth College.
http://www.dartmouth.edu/~rc/classes/data_management/s5.shtml

Appendix 2 - Glossary

Access – Continued, ongoing usability of a digital resource, retaining all qualities of authenticity, accuracy and functionality deemed to be essential for the purposes the digital material was created and/or acquired for.¹⁴

Archive – Place where objects are deposited with expectation that they may be accessed for use long into the future.

Authenticity – Promise that the digital object is complete and unaltered once it has been created. Metadata is used to establish authenticity.

Backup – Duplication of data either on-site or at a location removed from the original data. Assumes no managed activity to ensure data is accessible in the future.

¹¹ (Consultative Committee for Space Data Systems, 2002)

¹² (Library of Congress, 2008)

¹³ (OCLC; The Center for Research Libraries, 2007)

¹⁴ Digital Preservation Coalition, 2009

Born Digital – Digital materials which are not intended to have an analogue equivalent, either as the originating source or as a result of conversion to analogue form.²

Digital Preservation – The series of management policies and activities necessary to ensure the enduring usability, authenticity, discoverability, and accessibility of content over the very long term.”¹⁵

Digital Repository – A place where digital assets are deposited and stored.

File Format – An attribute of a file which describes its encoding.¹⁶ File formats are typically identified by a three or four letter extension at the end of a file name (i.e. .DOC, .MOV, .PDF, .XLS).

Life Cycle – A series of stages through which something, in this case digital information, passes during its lifetime. The lifecycle for digital information includes creation, use and reuse, migration or emulation, and storage.

Long-term Storage – A conscious decision to retain object in perpetuity or until agreements or selection policies change. Also implies management of object to migrate data as necessary to keep it accessible and understandable.

Metadata – A term that refers to structured data about data. "Preservation metadata" is the term for a broader set of metadata that documents the lifecycle of digital content from creation through processing, storage, preservation, and use over time.¹⁷

Migration – A means of overcoming technological obsolescence by transferring digital resources from one hardware/software generation to the next. The purpose of migration is to preserve the intellectual content of digital objects and to retain the ability for clients to retrieve, display, and otherwise use them in the face of constantly changing technology. Migration differs from the refreshing of storage media in that it is not always possible to make an exact digital copy or replicate original features and appearance and still maintain the compatibility of the resource with the new generation of technology.

Normalization – In a preservation context, normalization refers to a preservation strategy that involves the imposition of standard formats and rules to create preservable file formats. Normalization has specific connotations within the database (e.g., normalized tables), the Web (e.g., normalized URLs), and other communities, but the essence of the term is to standardize for more effective processing and exchange of information.⁵

¹⁵ (JISC, Digital Preservation Coalition, Digital Archives Department of the University of London Computer Centre, Portico, 2009)

¹⁶ (AHDS, 2003)

¹⁷ (ICPSR, 2007)

Appendix 3 - Bibliography

- AHDS. (2003, January). Retrieved December 13, 2010, from AHDS Digital Preservation Glossary: <http://www.ahds.ac.uk/preservation/preservation-glossary.pdf>
- Blue Ribbon Task Force on Sustainable Digital Preservation and Access. (2010, February). *Sustainable Economics for a Digital Planet: Ensuring Long-Term Access to Digital Information*. Retrieved January 3, 2011, from http://brtf.sdsc.edu/biblio/BRTF_Final_Report.pdf
- Checkley-Scott, C., & Thompson, D. (2007). *Wellcome Library Preservation Policy for Materials Held in Collections*. Retrieved January 3, 2011, from <http://library.wellcome.ac.uk/assets/wtx038065.pdf>
- Columbia University Libraries. (2006, July). *Policy for Preservation of Digital Resources*. Retrieved January 3, 2011, from Preservation and Digital Conversion Division: <http://www.columbia.edu/cu/lweb/services/preservation/dlpolicy.html>
- Consultative Committee for Space Data Systems. (2002, January). *Reference Model for an Open Archival Information System (OAIS)*. Retrieved January 10, 2011, from <http://public.ccsds.org/publications/archive/650x0b1.pdf>
- Cornell University Library; ICPSR. (2010, May). *Digital Preservation Management: Implementing Short-term Strategies for Long-term Problems*. Retrieved January 29, 2011, from http://www.icpsr.umich.edu/dpm/dpm-eng/eng_index.html
- Dartmouth College Library. (2007). *Self Study & Recommendations*. Retrieved January 10, 2011, from <https://www.dartmouth.edu/~library/home/staffweb/self-study/index.html>
- Dartmouth College Library. (2008). *Introduction to Records Management at Dartmouth College*. Retrieved January 31, 2011, from Records Management: <http://www.dartmouth.edu/~library/recmgmt/>
- Dartmouth College Library. (2009, August 21). *Library Mission and Goals FY2010*. Retrieved January 10, 2011, from <http://www.dartmouth.edu/~library/home/about/mission.html>
- Dartmouth College Library. (2010, November 18). *Library Mission and Goals FY2011*. Retrieved January 10, 2011, from https://www.dartmouth.edu/~library/home/staffweb/libonly/library_mission_and_goals_fy2011.html
- Dartmouth College Library. (2011). *Library Mission and Goals FY2012*. Retrieved August 8, 2011, from <https://www.dartmouth.edu/~library/home/about/mission.html>
- Digital Curation Centre. (2010). *DCC Curation Lifecycle Model*. Retrieved January 18, 2011, from <http://www.dcc.ac.uk/resources/curation-lifecycle-model>
- Digital Preservation Coalition. (2009). *Digital Preservation Handbook*. Retrieved December 13, 2010, from Definitions and Concepts: <http://www.dpconline.org/advice/preservationhandbook/introduction/definitions-and-concepts>
- Digital Projects and Infrastructure Group. (2010, April 27). *The Dartmouth Digital Library: Program, Priorities, and Policies*. Retrieved January 18, 2011, from Digital Projects and Infrastructure Group wiki: <https://libwiki.dartmouth.edu/twiki/pub/Libopen/DPIG/DDL-policy2.docx>

- ICPSR. (2007, June 18). Retrieved December 13, 2010, from Digital Preservation Glossary: <http://www.icpsr.umich.edu/icpsrweb/ICPSR/curation/preservation/glossary.jsp>
- JISC, Digital Preservation Coalition, Digital Archives Department of the University of London Computer Centre, Portico. (2009, April). Retrieved December 13, 2010, from JISC Project Report: Digitisation Programme:: http://www.jisc.ac.uk/media/documents/programmes/digitisation/jisc_dpp_final_public_report.pdf
- Library of Congress. (2008, March). *PREMIS Data Dictionary for Preservation Metadata version 2.0*. Retrieved January 10, 2011, from PREMIS Preservation Metadata Maintenance Activity: <http://www.loc.gov/standards/premis/v2/premis-2-0.pdf>
- National Library of Australia. (2008). *Digital Preservation Policy*, 3rd Edition. Retrieved January 3, 2011, from <http://www.nla.gov.au/policy/digpres.html>
- OCLC; The Center for Research Libraries;. (2007, February). *Trustworthy Repositories Audit & Certification: Criteria and Checklist*. Retrieved January 19, 2011, from http://www.crl.edu/sites/default/files/attachments/pages/trac_0.pdf
- University of Illinois at Urbana-Champaign. (2009, November). *IDEALS Digital Preservation Policy*. Retrieved January 3, 2011, from Illinois Digital Environment for Access to Learning and Scholarship: <https://services.ideals.illinois.edu/wiki/bin/view/IDEALS/IDEALSDigitalPreservationPolicy>
- Yale University Library. (2005). *Digital Preservation Policy*. Retrieved January 3, 2011, from <http://www.library.yale.edu/iac/DPC/final1.html>