Policies, Procedures, Guidelines

UNIVERSITY OF ALBERTA
Digital Preservation
http://www.library.ualberta.ca/digitization/preservation/

ASK US! chat phone email visit

UNIVERSITY OF ALBERTA LIBRARIES

MY ACCOUNT | LIBRARY HOURS | OUR LIBRARIES | version française

university of alberta

DI HOME | DIGITAL PRESERVATION | MORE DIGITAL SERVICES

Digital Preservation

University of Alberta Libraries has a strong commitment to the preservation of both analog and digital collections. As a major North American research library, we are advancing a number of digital preservation initiatives to ensure that our shared digital resources and collections are preserved for future generations of scholars, students, researchers and the general public.
We are seeking to create a Trusted Digital Repository within what we hope will be a trusted network of repositories across Canada.

> *"Has the Library been modernized? Have you thrown out the old tapes and computerizations? And always he imagined the answer from dusty and ancient librarians, 'As it has been, Professor, so is it still.'"*
>
> Isaac Asimov, Foundation's edge. (New York : Bantam, 1991, c1982), pp. 45-46.

INITIATIVES AND PARTNERSHIPS

1. **The Ebling Mis Initiative and the Sun Centre of Excellence**
   UAL is actively developing and implementing a scalable Trusted (Trustworthy) Digital Repository for the preservation of digital resources of various types. ERA (Educational Research Archive), digitized collections, audiovisual collections, data, purchased eBooks, and other materials are ingested (deposited) and safeguarded within this controlled environment so that they will continue to be accessible for generations of learners to come. For more information on some of the technology, metadata standards, and policies being employed in the preservation environment. Chief Librarian Ernie Ingles has dubbed this program The Ebling Mis Initiative, as borrowed from Asimov's novel Foundation's Edge.

   > The Library was outmoded and archaic - it had been so even in Ebling Mis's time - but that was all to the good. Pelorat always rubbed his hands with excitement when he thought of an old and outmoded Library. The older and the more outmoded, the more likely it was to have what he needed. In his dreams, he would enter the Library and ask in breathless alarm, "Has the Library been modernized? Have you thrown out the old tapes and computerizations?" And always he imagined the answer from dusty and ancient librarians, "As it has been, Professor, so is it still."
   > *Isaac Asimov, Foundation's edge. (New York : Bantam, 1991, c1982), pp. 45-46.*

   UAL is certainly not outmoded and archaic, but as Pelorat dreamed, it will do all that it must do to ensure the survivability of digital materials for generations to come

2. **Sun Preservation and Archiving Special Interest Group**
   The library is one of the original members of the Sun Preservation and Archiving Special Interest Group (Sun PA-SIG). The PA-SIG group is concerned with advancing best practices and practical implementation of the technologies supporting digital preservation.

3. **LOCKSS**
   The University of Alberta Libraries is a member of the worldwide LOCKSS Alliance and we actively preserve content from publishers that participate in this initiative.

4. **COPPUL PLN**
   UAL is a founding member of the COPPUL Private LOCKSS Network (PLN). The PLN is a pilot program that utilizes the LOCKSS digital preservation system as a means to archive collections of local interest to members of the Council of Prairie and Pacific University Libraries (COPPUL) that are not being preserved through any other means. Digital materials such as small university press publications, open access journals, born digital government publications, and other electronic journal collections that are at risk of being lost will be preserved as part of the program. This is the only LOCKSS PLN in Canada.

DOCUMENTATION

- Metadata Standards and Practices
- Technical Specifications

PEOPLE

- Peter Binkley
  Digital Initiatives Technology Librarian
- Robert Cole
  Bibliographer and Digital Content Coordinator, Peel's Prairie Provinces
- Sharon Farnel
  Metadata and Cataloguing Librarian
- Geoffrey Harder
  Digital Initiatives Coordinator
- Chuck Humphrey
  Data Library Coordinator
- Denise Koufogiannakis
  Collections and Acquisitions Coordinator
- Leah Vanderjagt
  Digital Repository Services Librarian

RELATED LINKS

- Canadian Digital Information Strategy
- Canadiana.org
- Know Alberta

5. **CLOCKSS**
   University of Alberta Libraries is an Archive Node of CLOCKSS, which involves members who act as custodians of the archived content. UAL is the only archive node in Canada. Chief Librarian and Vice-Provost, Learning Services, Ernie Ingles, is a member of the Board of Directors. More information about the CLOCKSS community is available on the CLOCKKS website.

6. **Portico**
   UAL is an Archive Founder member of Portico. As detailed on the Portico website, "Portico provides libraries and publishers with a reliable, cost-effective solution to one of the most critical challenges facing the scholarly community today - ensuring that the electronic resources you rely on everyday will be accessible to future researchers, scholars, and students.We preserve tens of thousands of e-journals, e-books, and d-collections (digitized historical collections) and are adding more content to the archive everyday. Libraries and publishers experience benefits from the Portico service from day one and can be confident that they will have access to their most valued content in the future."

7. **Web Archiving**
   University of Alberta Libraries has initiated a web archiving program in cooperation with the Internet Archive. We have licensed Archive-It to begin creating collections of important and strategic importance to the University of Alberta community.

   Collections include:

   Heritage Community Foundation
   University of Alberta Websites (in progress)
   Circumpolar websites (in progress)
   Western Canadian Political (in queue)

EXTERNAL DOCUMENTS AND REPORTS

Final Report of the Blue Ribbon Task Force on Sustainable Digital Preservation and Access
Trustworthy Repositories Audit & Certification: Criteria and Checklist (TRAC)
Ten Principles/Basic Characteristics of TDRs
Digital Curation and Preservation Bibliography by Charles W , Bailey, Jr.

ORGANIZATIONS

Canada
Canadian Association for Research Libraries (CARL) – Preservation
Canadian Association for Research Libraries (CARL) – Repositories
Council of Prairie and Pacific University Libraries (COPPUL)
Library and Archives Canada – Digital Preservation

International
Centre for Research Libraries – Archiving & Preservation
Digital Curation Centre (DCC)
Digital Preservation Europe
JISC
Library of Congress – Digital Preservation
Preservation and Archiving Special Interest Group(PA-SIG)

Columbia University *Libraries Digital Program*

Columbia's Long-Term Digital Preservation Archive (LTA)

Path: Digital Library Projects : Long Term Digital Preservation : Long Term Archive : **Status Report**

Columbia University Libraries/Information Services
Digital Preservation & Asset Management Infrastructure: Status Report

Columbia's Long-Term Digital Preservation Archive is a key component of Columbia's Digital Library and Institutional Repository infrastructure. It consists of a robust asset management system that can manage the digital resources of Columbia Libraries/Information Services for a variety of applications and at the same time provide the set of feature and services needed for long-term preservation of relevant digital assets. The need for a comprehensive architecture was identified when the new Digital Programs and Technology Services group (DPTS) was created in July 2007. Planning for the storage system began in the first quarter of 2008; implementation began in 2009.

Technical Architecture

The technical architecture has been designed with four main components:

1. Digital preservation storage system
2. Fedora software platform
3. Application and authentication middleware
4. Applications to support the Long Term Digital Preservation Achive and other programs

1. Digital Preservation Storage System

CUL/IS stores digital preservation assets on a total of four copies, two on disk and two on tape. One copy on disk and a second copy on tape will be located in an automated system in Columbia's main data center. A third copy on disk is located in the NYSERNet Data Center located in Syracuse, New York. A fourth copy on offline tape is sent to Iron Mountain to provide an additional offsite location.

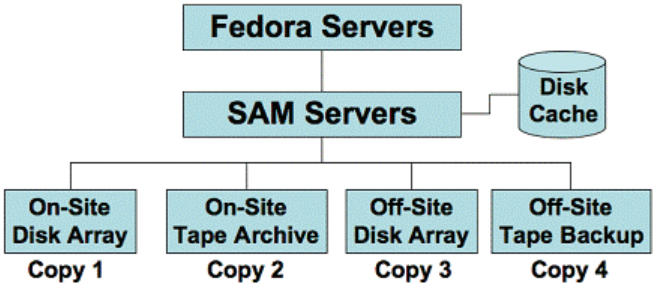
To manage multiple copies, automate migration and replication and provide a policy-based model to manage the long-term retention and access to digital assets, CUL/IS chose the Sun StorageTek Storage Archive Manager (SAM) software along with Sun hardware as a single vendor solution. SAM is "tried and true," with over a decade of proven use in managing large data repositories at corporations, supercomputer centers and libraries. It provides a self-protecting, automated data migration and recovery model that enables us to populate and incrementally expand the preservation storage to meet current and future needs. To support long-term sustainability and end-of-life data migration, SAM uses portable, nonproprietary data formats to store data on disk, the source code has been published as open source and uses open standards to provide data retrieval and access.

A total of 280 terabytes (TB) of disk and tape storage has been purchased to support the Digital Preservation Storage System. After this storage has been configured to support four copies of the digital assets, the system will have an effective storage capacity of approximately 70TB. As purchased, the system may be expanded incrementally to an effective storage capacity of up to 400TB. A high-speed, 10TB local disk cache provides increased access performance for commonly accessed digital assets and ensures that CUL/IS can rapidly load the system as required by large digital preservation efforts.

2. Fedora Software Platform

CUL/IS has chosen the Fedora Commons software platform to manage Columbia's digital repository, long term archive and a variety of other applications. Fedora version 3 has been installed on CUL/IS production servers, managed by Columbia University's central IT group. Fedora has been configured in

a "leader/follower" configuration, to provide replication and failover support.



3. Application and Authentication Middleware

CUL/IS has configured middleware to support common application needs, including but not limited to search and authentication/authorization. Authentication and authorization are built leveraging the University's identity management system that is based on Kerberos, LDAP and Shibboleth.

4. Long-Term Digital Archive Content and Applications

In 2009 we began loading content into our long-term digital preservation archive on the Fedora platform. As part of this ingest process, descriptive, structural and rights metadata will be reformulated according to current standards (e.g. MODS, METS, PREMIS, AES), and technical metadata will be generated automatically (e.g., using JHOVE) when feasible. New institutional policies and procedures required for meeting the standard of a "trusted digital repository" will be substantially complete by July 2011. The digital preservation repository will also be functionally integrated with our evolving digital asset management system, so that digital content of all kinds can be repurposed and for use by both the Columbia community as well as external users.

At present the range of content we will be preserving in our LTA include: digitized library collections (largely special and rare materials), content input into our local institutional repository (Academic Commons) and, on a test basis, web sites harvested as part of a grant-funded pilot program.

In implementing our local applications and policies supporting the LTA, we will be guided by the May 2002 report **"Trusted Digital Repositories, Attributes and Responsibilities: An RLG-OCLC Report"** along with other more recent research and best practices.

References

NYSERNet Data Center: http://www.nysernet.org/services/bcc/

Iron Mountain: http://www.ironmountain.com/dataprotection/vault/

Sun StorageTek SAM Software:
http://www.sun.com/storagetek/management_software/data_management/sam/

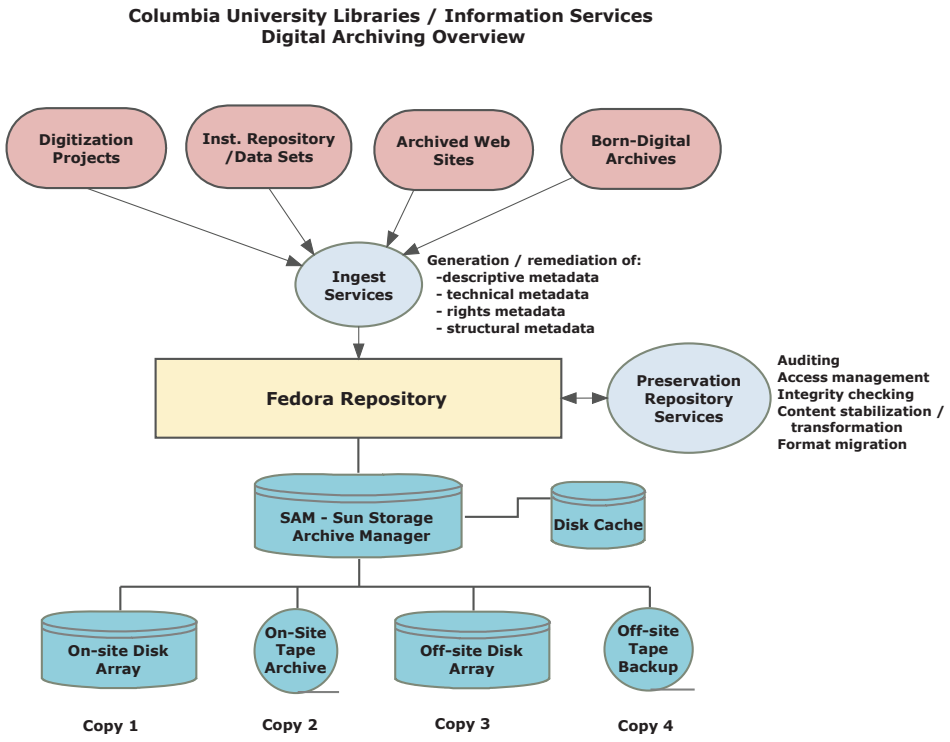
Columbia Libraries **Digital Program**

Last revision: 10/28/10
© Columbia University Libraries

COLUMBIA UNIVERSITY

Preservation Asset Overview

http://www.columbia.edu/cu/libraries/inside/projects/digital_pres/lta/preservation_asset_overview_current.pdf



UNIVERSITY OF FLORIDA
Florida Digital Archive
http://fclaweb.fcla.edu/FDA_documentation

FCLA

The Florida Center for Library Automation

GO

FCLA HOME | SUPPORT | ABOUT US | SYSTEMS STATUS | SERVICES | SUS LIBRARIES | CONTACT US

Home >

- Reports & Statistics
- Catalogs
- E-Resources
- Digital Collections
- Digital Preservation
- Acq/Cat/Circ/Etc.

The mission of the Florida Digital Archive is to provide a cost-effective, long-term preservation repository for digital materials in support of teaching and learning, scholarship, and research in the state of Florida.

FLORIDA Digital Archive

Digital Archive Information

Florida Digital Archive Terms

Library Agreement

- FCLA--Library Agreement (Word)
- Instructions for filling out Appendix A (PDF)

FDA Policy Guide

- Florida Digital Archive (FDA)
  Policy and Procedures Guide v.3.0 [May 2011]

Procedures and documentation for Florida Digital Archive Affiliates

- Ingest Report Warning Messages
- Archive Services Reports (Ingest and Error/Reject)
- METS DAITSS SIP Profile
- Florida Digital Archive (FDA) Sip Specification v.2.0 [May 2011]

FDA Report Stylesheet and Examples

FDA reports of package Ingest, Dissemination, Withdrawal, and Errors are XML documents. To view these reports, please download the FDA Report Stylesheet and save it to the same directory as your FDA reports.

The following are examples of each of the FDA reports:

- Ingest Report Sample 1
- Ingest Report Sample 2
- Error Report Sample 1
- Error Report Sample 2
- Dissemination Report Sample
- Withdrawal Report Sample

© 2009 The Florida Center for Library Automation, 5830 NW 39th Ave, Gainesville, FL 32606; 352.392.9020
Home | Contact Us| About This Site | This Site Uses Google Analytics (Google Privacy Policy) |

Florida Digital Archive

Florida Digital Archive (FDA) Policy and Procedures Guide

Version 3.0, May, 2011
Last reviewed May, 2010 without updates
superseded versions:
version 2.5, April 2009
version 2.4, August 2007
version 2.3, December 2006
version 2.2, August 2006
version 2.1, February 2006
version 2.0, January 2006

This document covers mission, governance, division of responsibilities, archivable materials, rights, services, preservation strategies and other topics of interest to FDA users.

© Copyright 2006 – 2011 by the Florida Center for Library Automation.

CONTENTS

Mission....3
Background....3
Governance....3
FDA Affiliates....4
Designated Community....4
Terminology: SIP, AIP and DIP....4
Model of operation and division of responsibilities....5
Archivable materials....7
Rights in archived materials....7
Billing....8
Services....8
Preservation Strategies....10
Storage....11
Security....12
Continuity of Operations Plan....13
Succession plan....13
Related documentation....13

Mission

The mission of the Florida Digital Archive (FDA) is to provide a cost-effective, long-term preservation repository for digital materials in support of teaching and learning, scholarship, and research in the state of Florida.

In support of this mission, the FDA guarantees that all files deposited by agreement with its Affiliates remain available, unaltered, and readable from media. For supported formats, the FDA will maintain a usable version using the best format migration tools available.

Background

Planning for the FDA began in 2001 in response to the perceived need of the directors of the libraries of the public universities of Florida to ensure the permanent availability of digital library materials such as electronic dissertations. Development was expedited by the award in 2002 of a three-year grant from the Institute of Museum and Library Services (IMLS) which concluded in September, 2005.

In order to implement the FDA, FCLA staff designed and developed the DAITSS application (Dark Archive in the Sunshine State). In November, 2005, an early version of the DAITSS application that lacked dissemination and withdrawal functions went into production for the Florida Digital Archive. In December 2006 the first version of DAITSS with all major planned functionality was completed. The software was released as open source under the GPL license in 2007. In April 2011 a wholly rearchitected and recoded version of DAITSS was installed as DAITSS 2.

The technical design, procedures and policies of the FDA are based on OAIS - Open Archival Information System Reference Model (ISO 14721:2003) and on ongoing work to define and certify trusted digital repositories, including *Trusted Digital Repositories: Attributes and Responsibilities* (RLG May 2002), the RLG/NARA *Audit Checklist for Certifying Digital Repositories* (RLG August 2005), and *Trustworthy Repositories Audit & Certification: Criteria and Checklist* (NARA, *et al.*, February 2007).

Governance

The FDA is run by the Florida Center for Library Automation, a system-wide center of the state universities attached to the University of Florida for administrative purposes. It is under the administrative control of the Director of FCLA.

The FCLA Advisory Board acts as the Advisory Board of the FDA, recommending policy decisions to the FCLA Director. The FCLA Advisory Board

updated AIP is what is returned to the Affiliate as a DIP. This process ensures that the disseminated package is always as complete and up-to-date as possible.

5. Withdrawal.

(Withdrawal is temporarily unavailable in DAITSS 2.0.) An AIP will be withdrawn from the FDA upon the request of an authorized agent of the Affiliate. All withdrawal requests must be sent to the FDA Help Desk at FDA@prb.fcla.edu. Files belonging to the withdrawn AIP are deleted entirely from storage, but the FDA retains a permanent record that the intellectual entity was ingested and withdrawn. The Affiliate will receive a Withdrawal Report by email.

A common reason for withdrawal is to update information in a previously submitted package. The existing AIP must be withdrawn, and a new SIP must be submitted for Ingest.

An AIP may be withdrawn on the initiative of the FDA, if the FDA receives information that the AIP has been archived in violation of copyright.

6. Reporting.

The FDA provides periodic statistical reports to Affiliates about their own use of the FDA. General statistical reports are posted to the FDA web site. *Ad hoc* reporting is also available upon request.

Preservation Strategies

Preservation strategies supported by the FDA are based on format transformation, that is, changing file formats to delay or accommodate format obsolescence. The FDA performs two kinds of transformations:

- *Normalization*. If a file is in a format considered to be less than optimal for digital preservation a version of the file may be created in a more preservation-worthy format. In general, preferred formats are non-proprietary, well documented, and well understood by FDA staff. Normalized versions may not be equivalent to originals in appearance or functionality. For example, a PDF file (WAV example) might be normalized into a set of page-image TIFFs. In this case the appearance of the content is retained but functionality such as actionable hyperlinks is lost. If normalization is part of the Action Plan for a particular file format, files in that format will be normalized on ingest. The normalized version will not be stored in the AIP, but the entire SIP will be rejected if normalization fails. This ensures that normalization can be done if necessary, but spares the cost of storing normalized versions.

- *Migration*. If a file is in a format considered at risk of obsolescence, a version may be created in a format considered to be a reasonable successor to the original format. All effort will be made to retain the appearance and behaviors of the original version, although this can not always be guaranteed. The successor format may be a higher version of the original format (for example, PDF 1.4 might be migrated to PDF 1.6) or it may be another format. If migration is part of the Action Plan for a particular file format, files in that format will be migrated on ingest.

The preservation strategies that will be implemented for any file format are documented in the Action Plan for the file format, available on the FDA website. Action Plans are reviewed periodically and revised when appropriate.

All preservation strategies are applied at the time a SIP is ingested, as part of ingest processing. This includes packages that are disseminated and then re-ingested, either as part of the Archive's planned preservation processes or as part of an Affiliate-requested dissemination (see next paragraph). Normalized and migrated versions of files contained in the SIP become part of the AIP.

If there is not an implemented Action Plan for the file format, bit-level preservation will be carried out for the file until the time when full preservation becomes available. At that time, the AIP containing the file can be disseminated and re-ingested, causing the full preservation treatment to be applied.

Storage

For every file in the AIP, two master copies are written. One copy is stored at the UF Computing & Network Services facility in Gainesville (CNS) and one copy is stored at the Northwest Regional Data Center in Tallahassee (NWRDC).

The two master copies are treated as a single file by DAITSS, the repository software application underlying the FDA. This means that when any action is performed on a file, it must be successfully performed on both master copies to be considered complete. For example, a fixity check involves calculating a message digest over the bits of a file and comparing this to a previously stored message digest. For a fixity check to be complete, message digests must be calculated for both of the master copies of the file and verified to match the stored message digest.

In addition to the master copies, traditional backup copies on tape are maintained in Gainesville and Tallahassee.

Security

Data security is ensured by a combination of physical security and cybersecurity.

SMARTech

SMARTech, or Scholarly Materials And Research @ Georgia Tech, is a repository for the capture of the intellectual output of the Institute in support of its teaching and research missions. SMARTech connects stockpiles of digital materials currently in existence throughout campus to create a cohesive, useful, sustainable repository available to Georgia Tech and the world.

See the Mission and Collection Policy .

Why should I participate?

- Access barriers disappear
- Enhanced visibility, use, reputation
- Wide and rapid dissemination of intellectual output
- Supports classroom teaching
- Aids multidisciplinary inquiry
- Valuable recruiting tool
- Preservation and management of information assets
- Reduces duplication of effort
- Stimulates serendipitous discovery and collaboration

What types of materials can I submit and find in SMARTech?

SMARTech houses Georgia Tech research in digital format, including

- Annual Reports
- Conference Papers
- Electronic Theses & Dissertations
- Learning Objects
- Newsletters
- Pre-Prints/Post-Prints
- Proceedings
- Research Reports
- Simulations
- Technical Reports
- Web Pages
- White papers
- Working Papers

What file formats are accepted?

We accept standard formats that we can make a commitment to migrate and provide access to over the long term including:

Type	Description	File extension	Support level
Text/Images	Adobe PDF	pdf	supported
Text	HTML	htm, html	supported
Text	Rich Text Format	rtf	supported
Text	Text	txt	supported
Text	XML	xml	supported
Text	Microsoft Word	doc	known
Text	WordPerfect	wpd	known
Text	SGML	sgm, sgml	known

Images	JPEG	jpg, jpeg	supported
Images	GIF	gif	supported
Images	PNG	png	supported
Images	TIFF	tif, tiff	supported
Images	Post Script	ps, eps, ai	supported
Images	BMP	bmp	known
Images	Adobe Photoshop	pdd, psd	known
Images	Microsoft Powerpoint	ppt	known
Images	Photo CD	pcd	known
Video	MPEG	mpg, mpeg, mpe	supported
Video	Video Quicktime	mov, qt	known
Audio	WAV	wav	supported
Audio	MPEG	mpa, abs, mpeg	supported
Audio	AIFF	aiff, aif, aifc	supported
Audio	RealAudio	ra, ram	known
Audio	Basic	au, snd	known
Special	Microsoft Excel	xls	known
Special	Microsoft Project	mpp, mpx, mpd	known
Special	Microsoft Visio	vsd	known
Special	FileMaker/FMP3	fm	known
Special	LateX	latex	known
Special	Mathematica	ma	known
Special	Tex	tex	known
Special	TeXdvi	dvi	known

supported Items in this category can be used in the future through migration or emulation and the Library makes a commitment to do so.

known This category indicates that the specifics of the program code for that format are not public but the format is so widely used that the ability to use it in the future is almost certain.

How are materials in SMARTech preserved?

SMARTech is part of the MetaArchive Cooperative distributed digital preservation network. Georgia Tech Library participates in the MetaArchive program, an international effort for the preservation of electronic scholarly materials through the Library of Congress' National Digital Information Infrastructure and Preservation Program (NDIIPP).

How do I start contributing to SMARTech?

- **email:** smartech@library.gatech.edu

Mission and Collection Policy

Statement of Purpose

The Scholarly Communication and Digital Services Department is charged with building effective, dynamic knowledge management and research systems to preserve and provide access to the intellectual output of Georgia Tech. Its responsibilities include identifying, assessing, collecting, preserving, providing access, and making this output more valuable through digital information technologies, whether "born-digital" or convertible to digital formats. The Department provides these resources and services in support of the research and educational endeavors of the Georgia Tech community and to scholars around the world.

Collecting Priorities: Subject Areas and Forms of Intellectual Output

The Scholarly Communication and Digital Services Department's collecting areas reflect the major areas of research of the Georgia Tech campus, especially Science and Engineering. Effort is concentrated on those areas with national rankings and emerging significance. Materials range from traditional scholarly publications through materials created in support of classroom teaching. Through survey of the Institute's Web presence, personal interviews, or contact, materials are identified, investigated in light of content, technical, and legal considerations, and solicited/collected from the departments.

Centers of Excellence

U.S. News & World Report 2006 America's Best Colleges and America's Best Graduate Schools editions provides us with a ranking guide to Georgia Tech's exceptional research and educational centers on campus. This, used in conjunction with patent awards and dollar amounts of sponsored research projects leads us to our primary collecting areas.

Engineering

- Aerospace/Aeronautical/Astronomical
- Biomedical/Bioengineering
- Civil
- Computer Engineering
- Electrical/Electronic/Communications
- Environmental/Environmental Health
- Industrial/Manufacturing
- Computer Engineering
- Mechanical

Public Affairs

- Information and Technology Management

Science

- Chemistry
- Computer Science
- Computer Science Systems
- Applied Mathematics
- Mathematics
- Physics

Business

Areas recognized as Research Initiatives for an Enhanced Research Enterprise at Georgia Tech : [1]

- Microelectronics
- Nanoscience and technology
- Bioscience and technology
- Manufacturing

- Entrepreneurship
- Sustainability
- Telecommunications

Materials of specific interest:

- Conference Papers
- Data Sets
- Electronic Theses and Dissertations
- Learning Objects/Instructional Materials
- Pre-Prints/Post-Prints
- Proceedings
- Recorded Lectures and Symposia
- Research and Technical Reports
- Simulations
- Web Pages
- White papers
- Working Papers

Electronic Publications

In concert with the Archives, SMARTech collects electronic publications and traditional archival materials now available in digital format. This continues the Archives' mission of collecting and preserving the history of campus, promoting research and scholarship through collections relating to the academic curriculum, provides research experience for students in the use of primary sources, and aids in preservation of the legal and administrative documents of the Institute. Some of the materials collected by Archives for inclusion in SMARTech include:

- Newsletters
- Annual Reports
- Campus Radio Programs
- Speeches
- Lecture Series and Symposia

Assessment Criteria

All materials considered for inclusion in SMARTech will be assessed using the following criteria.

Content Considerations:

In addition to the subject areas for collection outlined above, materials will be assessed on their enduring value.

Technical Considerations:

Both the Library's ability to commit to the preservation of the bitstream/digital item and the technical quality of the materials will be considered in accepting materials for inclusion in SMARTech. Digital file formats are discussed below.

Legal Considerations:

The Library will research publishers' restrictions and conditions for deposit of pre-/post-prints, conference proceedings, and other items and ensure there is evidence of clear ability under copyright for the depositor to post the material.

Digital File Formats

Items considered for SMARTech will be evaluated using the same generally accepted categories standardized by MIT. [2]

Supported

When an item's format is public and open as is the case with formats such as Adobe PDF, HTML, JPEG, or AIFF it is categorized as a "Supported format." Items in this category can be used in the future through migration or emulation and the Library makes a commitment to do so.

- Adobe PDF (pdf)
- XML (xml)
- HTML (html, htm)
- Rich Text (rtf)
- Text (txt)
- Post Script (ps, eps, ai)
- GIF (gif)
- PNG (png)
- JPEG (jpg, jpeg)
- TIFF (tif, tiff)
- WAV (wav)
- MPEG (mpa, abs, mpeg)
- AIFF (aiff, aif, aifc)

Known

When an item is submitted in a proprietary format it is categorized as "Known." This category indicates that the specifics of the program code for that format are not public but the format is so widely used that the ability to use it in the future is almost certain.

- RealAudio (ra, ram)
- Basic (au, snd)
- Microsoft Excel (xls)
- Microsoft Project (mpp, mpx, mpd)
- Microsoft Visio (vsd)
- FileMaker/FMP3 (fm)
- LateX (latex)
- Mathematica (ma)
- Tex (tex)
- TeXdvi (dvi)
- Video Quicktime (mov, qt)
- BMP (bmp)
- Adobe Photoshop (pdd, psd)
- Microsoft Powerpoint (ppt)
- Photo CD (pcd)
- Microsoft Word (doc)
- WordPerfect (wpd)
- SGML (sgml)

Unsupported

"Unsupported" formats are those that the Library can not commit to converting to some usable form in the future. In consultation with the depositor, a decision will be made as to including the item in SMARTech and if it is accepted, readable descriptive information will be included. In the case of unsupported formats, the Library will request that the item also be submitted in a supported or known format, if it is at all possible to do so.

Procedures

Materials can be deposited into SMARTech in two main ways. The Library can receive permission from the submitter to deposit the material for them. The submitter is also able to deposit materials themselves: training from the library is available.

Withdrawal

Materials in SMARTech are considered permanent in nature. However, items raising legal issues, containing plagiarized content, or considered breaches of confidentiality may be considered for withdrawal. To avoid loss of

the historical record, any withdrawal transactions will be traced in the form of a note in the <Description.provenance> field of the Dublin Core record.

The content of the note should be one of the following:

"removed from view at request of the author"
"removed from view by legal order"
"removed from view at Georgia Institute of Technology's discretion"
"removed from view at Georgia Institute of Technology Library and Information Center's discretion"

Since any SMARTech item that has existed at some time may have been cited, we will always supply a "tombstone" when the item is requested, which will include the original metadata (for verification) plus one of the above withdrawal statements in place of the link to the object. For preservation and reference purposes, the item itself will be placed in Aardvark, a dark repository managed by the Georgia Tech Archives.

Access

Generally, items in SMARTech are open to the public. However, there are instances where copyright law or specific needs of the depositor requires campus-only access. The Library always encourages open access to all items found in SMARTech but will review requests for limited, campus-only access on a case by case basis.

Preservation

The Library is committed to adhering to the best practices of the profession applying to digital preservation. All materials are backed up with at least one method. All servers are backed up by tape. The Library is a member of the MetaArchive cooperative, a Digital Preservation Partner of the library of Congress' National Digital Information Infrastructure Preservation Program (NDIIPP) for the networked preservation of cultural heritage materials using the LOCKSS software. All Georgia Tech created digital materials held by the Library are included in this network. The Library also participates in the ASERL LOCKSS program for the preservation of electronic theses and dissertations. In addition, dissertations are submitted to UMI/ProQuest which is the official off-site repository for the preservation of dissertations for the Library of Congress.

This mission and collecting policy will be reviewed and revised based upon the experience gained in implementing the policy. A review and assessment of the policy will be conducted every 6 months by the Scholarly Communication and Digital Services staff and others invited to participate in the review.

[1] http://www.gatech.edu/president/strategicplan.html

[2] http://libraries.mit.edu/dspace-mit/build/policies/format.html

UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN
IDEALS Policies and Guidelines
https://services.ideals.illinois.edu/wiki/bin/view/IDEALS/PolicyAndGuideDocs

IDEALS

Jump | Search

- IDEALS Home
- IDEALS Resources

Using IDEALS
- Getting Started
- FAQ
- IDEALS Help
- Policies and Guidelines
- Deposit Agreement

Reports
- Monthly Downloads

More Information
- About IDEALS
- Presentations and Publications
- Contact IDEALS

- Site Map
- Search

- Administrative Resources
- *Staff Only Area*

- Log In

IDEALS > Wiki Home > > IDEALS Web > PolicyAndGuideDocs (28 May 2010, SarahShreeves) | Edit | Atta

IDEALS Policies and Guidelines

This page details all current IDEALS Policies and Guidelines.

Deposit Policies and Guidelines

General Deposit Guidelines for IDEALS: General deposit guidelines for IDEALS

Preparing Files for Deposit into IDEALS: Guidelines for preparing files for deposit into IDEALS.

Frequently Asked Questions: Frequently asked questions from people depositing content into IDEALS.

Individual Depositor Policies: The roles, rights, and responsibilities of individuals depositing research into the UIUC Researchers and Scholars community.

Community Policy: The roles, rights and responsibilities of IDEALS Communities.

Community Liaison/Admin FAQ: Frequently asked questions from Community Liaisons or Community Administrators.

Content Policies and Guidelines

Collection Policy: Describes the type of content that is acceptable for submission into IDEALS.

Access Restriction Policy: Describes how IDEALS deals with submissions which require restricted access, and the current levels of access available within IDEALS

Withdrawal Policy: Describes what occurs when someone requests that a submission be removed from the IDEALS repository.

Metadata Policy: Describes required and optional fields for describing documents submitted into IDEALS.

Metadata Guidelines for Deposited Items: A guide to editing metadata for items that have been deposited.

Dissertations and Theses Policies and Guidelines: Policies and Guidelines for Dissertations and Theses in IDEALS

Digital Preservation Policies

Digital Preservation Policy: The goals and operating principles of IDEALS in terms of preserving submitted digital content.

Digital Preservation Support Policy: Describes the "categories" of preservation support which IDEALS promises for different file formats.

Format Recommendations: Provides recommendations for file formats to deposit with IDEALS.

Copyright and Intellectual Property Policies

Copyright and Intellectual Property Policy: Details whether you will be able to deposit your research into IDEALS based on who owns the copyright or intellectual property rights.

Deposit Agreement: Non-Exclusive Distribution and Preservation License: The license agreement displayed to the submitter for each item submitted to IDEALS.

Current Publisher Policies on Archiving in Institutional Repositories: A quick list of publishers whose archiving policies are well known, compiled for reference by IDEALS. This list is meant as a compliment to the Sherpa/RoMEO database, which has a much larger listing of such publisher policies.

Support, Access, and Use Policies

Service Level Definition: A description of the IDEALS service.

Access and Use Policy: A description of the access and use policies for IDEALS.

Pilot Phase Support and Back-Up Policies - No Longer Applicable - see Service Level Definition: What is supported during the pilot phase of IDEALS.

Edit Attach Print version History: r16 < r15 < r14 < r13 < r12 Backlinks Raw View WYSIWYG edit More topic actions

Topic revision: r16 - 28 May 2010 - 10:51:04 - SarahShreeves

Copyright © 2011 by University of Illinois at Urbana-Champaign
Il material on this collaboration platform is the property of the University of Illinois at Urbana-Champaign.
uestions, requests, or problems finding IDEALS Resources? Send feedback

IDEALS Digital Preservation Policy

(November 2009)

Introduction

Preservation, as it applies to library and archives material, can be defined as: "all managerial and financial considerations including storage and accommodation, provision, staffing levels, policies, techniques and methods involved in preserving library and archive materials and the information contained therein." (Mirjam M. Foot, 'Building Blocks for a Preservation Policy.' (London: The National Preservation Office, 2001): 1.)

Within the context of an institutional repository, these combined activities allow the University to better ensure access to digital scholarly content for as long as possible by whoever needs it. Committed to building collections of digital resources and maintaining them for use by students, faculty, scholars, and the public long into the future, the University of Illinois assumes a long term stewardship obligation for the content within the institutional repository, the Illinois Digital Environment for Access to Learning and Scholarship (IDEALS).

Purpose

The IDEALS Digital Preservation Policy outlines this program's plan to support sustainable access to deposited content for the foreseeable future. This policy is subject to change as new and emerging technologies impact our ability to preserve deposited content. However, the development of a reliable digital archive that adheres to and remains compliant with changing standards and best practices remains the best opportunity for IDEALS' success as a repository for the research and scholarship of the University of Illinois.

Mandate

The IDEALS mandate for ensuring the preservation of deposited digital content stems from three primary areas of responsibility:

- Preservation of Institutional Scholarship: As an institution of higher education, the University of Illinois is obligated to support scholarship, teaching, and learning. As a repository for locally produced digital research and scholarship, IDEALS responsibilities include the identification, stewardship, and preservation of deposited content.
- Contractual and Legal Obligation: To the extent that deposited content requires persistent, contractually mandated legally-binding access, IDEALS is mandated to preserve the content.
- Organizational Commitment: The IDEALS partnership and funding is predicated upon the institution's obligation and commitment to developing a repository that provides access and preserves locally developed digital research and scholarship

Sarah L. Shreeves *IDEALS* *Last Updated Nov 2009*

Objectives

IDEALS' objectives are to collect, preserve, and maintain access to significant components of the University's scholarly and research output. IDEALS shall:

- Provide a reliable preservation environment for digital research and scholarship developed at the University of Illinois.
- Provide reliable and consistent access to all digital research and scholarship entrusted into its service.
- Build a trusted service for scholars at the University of Illinois.

Scope

The IDEALS initiative is responsible for preserving, managing, and providing access to locally developed and deposited digital research and scholarship. In particular, it is responsible for finished research and scholarly content and associated supporting documentation and data.

Operating Principles

The IDEALS initiative shall adhere to the following operating principles:

- IDEALS strives to comply with the Open Archival Information System (OAIS) Reference Model standard.
- IDEALS strives to comply with certification requirements for a Trusted Digital Repository.
- IDEALS adheres to prevailing community standards for preserving access to digital content whenever possible.
- IDEALS participates in the development and implementation of standards.
- IDEALS commits to an interoperable, scalable digital archive with appropriate storage management for content.
- IDEALS policies, procedures, and practices are clearly documented and consistent.
- IDEALS maintains hardware, software, and storage media containing archival content in keeping with prevailing best practices.
- IDEALS establishes procedures to meet archival requirements pertaining to provenance, chain of custody, authenticity, and integrity.
- IDEALS complies with intellectual property, copyright, and ownership rights for all content.

Sarah L. Shreeves *IDEALS* *Last Updated Nov 2009*

Roles and Responsibilities

Libraries and archives have long been entrusted with ensuring access to scholarly content. As a joint program operated between the University Library and CITES on the University of Illinois at Urbana-Champaign campus, the IDEALS initiative seeks to bridge two very different operational models. Yet, the primary objective remains the same - to preserve and provide access to digital research and scholarship.

Both the University Library and CITES assume joint responsibility for the long-term preservation of and access to deposited content.

Selection and Acquisition

The collection will focus on deposits of digital research and scholarship by faculty, students and staff of the University of Illinois. Contributors may include non-affiliated scholars if they are co-authoring with University of Illinois authors or are affiliated closely with the University, e.g., are emeritus professors, Survey personnel, or hold honorary appointments. See the IDEALS Collection policy for more information.

Preservation Strategies

Digital preservation management activities in the IDEALS initiative include:

- Development and maintenance of reliable options for the ingest of new materials into the repository, based on community standards or best practices;
- Provision of reliable data management services for timely access to deposited content;
- Development and maintenance of archival storage for deposited content;
- Conducting IDEALS management and administrative activities in such a manner as to further the program's mission of preserving deposited content;
- Monitoring and remaining active in community preservation activities, best practices and standards; and
- Developing local preservation planning activities that will anticipate and respond to changes in the preservation environment (e.g. format migration or emulation strategies).

Access/Use

By default all materials deposited into IDEALS are available to any user with internet access. However in some cases, access restrictions may be necessary to comply with intellectual property or copyright agreements (i.e. publisher embargoes). The expectation is that materials accepted for deposit will be renderable (i.e. human viewable) according to the limitations and opportunities of current technologies deployed by IDEALS. This means there may be some format types (usually obsolete or obscure) that will

not be fully renderable and when this is the case, IDEALS will make such limitations known in advance whenever possible.

Use of materials is limited by whatever copyright notice has been provided. In most cases, authors reserve all rights and material can only be used under fair use provisions.

See the Access and Use Policy for more information.

Challenges

- Technological Change: Developing a sustainable digital preservation model that will respond to technological changes as needed without under- or overestimating the needs imposed by these changes.
- Creation of a Development Environment: Developing an OAIS compliant model requires a preservation planning environment that will permit program personnel to experiment with technological and procedural changes without risk of damaging deposited content.
- Sustainability: Developing a sustainable model that will deal with the technical and management challenges of preserving born-digital content within the constraints of available funding.
- Full Engagement in Preservation: Providing a thoughtful balance between access and preservation while being mindful of preservation's core role in maintaining access.
- Sustaining the Relationship between the University Library and CITES: Maintaining IDEALS is the joint responsibility of the University Library and CITES. Defining, developing, and sustaining this unique relationship between the Library and CITES will ensure this program's long term stability and success.
- Ongoing Monitoring of Submitted Material Types and Formats: As different type of materials are submitted (data sets, complex information objects), monitoring different needs (storage size, metadata, etc) of the materials and maintaining procedures and policies (i.e. Format Support Policy) based on these needs is necessary.

Cooperation/Collaboration

The University Library and CITES are committed to collaborating with one another in the development of a Trusted Digital Repository in IDEALS. The IDEALS initiative is also committed to collaborating with other institutions and organizations to further technological and operational research to better serve the collective desire to digital content.

Sarah L. Shreeves *IDEALS* *Last Updated Nov 2009*

IDEALS | Illinois Digital Environment for Access to Learning and Scholarship

Jump | Search

- IDEALS Home
- IDEALS Resources

Using IDEALS
- Getting Started
- FAQ
- IDEALS Help
- Policies and Guidelines
- Deposit Agreement

Reports
- Monthly Downloads

More Information
- About IDEALS
- Presentations and Publications
- Contact IDEALS

- Site Map
- Search

- Administrative Resources
- *Staff Only Area*

Log In

IDEALS > Wiki Home > > IDEALS
Web > PolicyAndGuideDocs > PreservationSupportPolicy (07 Feb 2008, kcostel4)

Edit | Attach

IDEALS Digital Preservation Support Policy

Committed to building and maintaining collections for the use of students, faculty, scholars, and the public long into the future, the University of Illinois at Urbana-Champaign assumes an obligation to ensure long-term access to the materials deposited into IDEALS and their intellectual content, but also acknowledges the inherent challenges involved in preserving digital content.

To this end, the IDEALS Digital Preservation Support Policy defines the categories of preservation support available and provides specific information about where different file formats fit within these categories. This policy is subject to change as new and emerging technologies impact our ability to preserve deposited content.

Table of Contents

- IDEALS Digital Preservation Support Policy
  - Background
  - Categories of Preservation Support
    - Category 1 - Highest Confidence - Full Support
    - Category 2 - Moderate Confidence - Intermediate Support
    - Category 3 - Low Confidence - Basic Preservation Only
  - Table of Preservation Actions

Background

Our ability to preserve digital objects deposited in IDEALS is dependent, among other things, on whether the file format used:

- Is openly documented (more preservable) or proprietary (less preservable);
- Is supported by a range of software platforms (more preservable) or by only one (less preservable);
- Is widely adopted (more preservable) or has low use (less preservable);
- Is lossless data compression (more preservable) or lossy data compression(less preservable); and
- Contains embedded files or embedded programs/scripts, like macros (less preservable).

All digital objects deposited to IDEALS will receive a basic level of preservation. *Basic preservation* means that IDEALS will preserve the *viability* of the original

object through:

- ensuring that the bitstream (the 1s and 0s that make up the digital file) remains exactly the same over time;
- assigning a persistent, permanent identifier;
- creating preservation metadata;
- maintaining onsite and offsite backup copies;
- performing regular virus and file corruption checks; and
- performing periodic refreshments by copying files to new storage media.

Basic preservation does **not** ensure that a digital object may be opened by a computer program or is understandable by a human in the future. For example, in 2006 a faculty member deposits a conference presentation in the Microsoft PowerPoint format (ppt), a proprietary format. In 2030, a graduate student would like to view that conference presentation, but the software program - Microsoft PowerPoint - used to open and read ppt files has been discontinued since 2020. Old versions of the software program are difficult to find, and, because the ppt file format had never been publicly documented, there exist no other software programs to open the file. Even though the original digital object (the conference presentation in ppt) is still technically *viable*, it is no longer *renderable* (able to be opened by a computer program), and thus not *understandable* by the graduate student in 2030.

Therefore, for digital objects that meet certain criteria (see below), IDEALS will strive to preserve not only the *viability* of the object but also the *renderability* and the *understandability* of the content of the digital object, as well as the original file itself. In the case of some objects in proprietary formats, this will mean that in addition to the original digital object, IDEALS will also save a copy of the object transformed into a file format that is more preservable than the original. For example, the conference presentation in ppt might also be saved as a pdf/a object (an open, publicly documented standard). The pdf/a object is a more preservable format than the ppt format. What may be lost is the full *functionality* of the original digital object. For example, the graduate student in our example may not be able to view the conference presentation as a slide show as the Microsoft PowerPoint software program allows. However, the content of the conference presentation will be preserved.

IDEALS also recognizes that in some cases an *access copy* of a digital object is necessary due to the proprietary nature or cost of the software used to render it. For example, a Microsoft Word document is reliant on the Microsoft Word progam to render it; IDEALS will also provide a pdf version of the document because pdf readers are freely and readily available. In some cases, the *access copy* and the *preservable copy* may be the one and the same - a pdf/a version, for example.

Categories of Preservation Support

IDEALS categorizes digital objects into three categories of preservation support. These categories are defined below. Any format not yet reviewed and evaluated by IDEALS will receive Category 3 support on deposit. A different category may be assigned after format review takes place.

Category 1 - Highest Confidence - Full Support

Description:

- Most confidence in ability to provide long term preservation to **content** and **functionality**
- Highest level of preservation support in effort to maintain *viability*, *renderability*, and *understandability* as well as *functionality* of original digital object.

Criteria:

- Is in a format this is publicly documented (example: xml);
- Is in a format this is widely adopted (example: xhtml);
- Is in a format that may be rendered by multiple software packages (example: txt);
- Is in a format that has lossless data compression (example: uncompressed tiff files); and
- Contains no embedded files or dynamic content (example: txt).

Actions:

- Monitor file format for changes that might warrant transformation or reassessment;
- Migration of document to successive format when necessary;
- Basic preservation including:
    - bitstream maintenance;
    - persistent, permanent identifier;
    - preservation metadata;
    - onsite and offsite backup copies;
    - regular virus and file corruption checks;
    - periodic refreshments to new storage media.

Examples:

- Plain text document in unicode
- A tiff image

Category 2 - Moderate Confidence - Intermediate Support

Description:

- Moderate confidence level in ability to provide long term preservation to **content** of file
- Intermediate level of preservation support in effort to maintain maintain *viability*, *renderability*, and *understandability* (but not functionality) of original digital object.

Criteria:

- Is in a format that is publicly documented;

- **AND** is in a format that has lossy data compression (example: Ogg Vorbis);
- **OR** is in a version of a format that has been deprecated in favor of a later version (example: HTML 3.0).

OR

- Is in a proprietary format;
- Is in a format that is widely adopted; and
- Is in a format that is of enough public and/or commercial interest that tools are likely to be available to migrate them to successor formats.

NOTE: Files with embedded content (for example, a PowerPoint (ppt) with a AVI video file (avi) inserted into it) are more preservable if the the files are deposited as separate files within the same item in IDEALS `need link here to help screen`. If the content remains embedded, it will likely not remain intact when the file is transformed to a more preservable format.

NOTE: Files with dynamic content (for example, an Excel spreadsheet (xls) with dynamic functions - even simple ones!) are more preservable if the dynamic content is either documented (for example, a note in an Excel spreadsheet explaining the functions that are included) or the document is saved as a static document (for example, a cell in an Excel spreadsheet that is the sum of a column is saved as the sum, not the function of adding the multiple cells).

Actions:

- Monitor file format for changes that might warrant transformation or reassessment;
- When possible, transformation to a format that preserves the *content* and when possible the *formatting and style* of the original, but not necessarily the *functionality*.
- Basic preservation of original object including:
  - bitstream maintenance;
  - persistent, permanent identifier;
  - preservation metadata;
  - onsite and offsite backup copies;
  - regular virus and file corruption checks;
  - periodic refreshments to new storage media.

Examples:

- Microsoft Word document (proprietary format)
- A compressed TIFF file.

Category 3 - Low Confidence - Basic Preservation Only

Description:

- Low confidence level in ability to provide long term preservation to **content** of file
- Basic level of preservation support in effort to maintain maintain *viability* of

original digital object **only**.

Criteria:

- Is in a proprietary format;
- Is in a format about which little information is publicly available;
- Is in a format that is not widely adopted;
- Is in a format with lossy data compression;
- Is supported by a single or very few software platforms; and/or
- Is in a format that does not meet the criteria for any of Categories 1-2.

Actions:

- Basic preservation of original object only including:
  - bitstream maintenance;
  - persistent, permanent identifier;
  - preservation metadata;
  - onsite and offsite backup copies;
  - regular virus and file corruption checks;
  - periodic refreshments to new storage media.

Examples:

- Kodak Photo CD format (pcd)
- Windows Media Video (wmv)

Table of Preservation Actions

Preservation Action	Category 1	Category 2	Category 3
Provision of persistent identifier for object and/or its metadata	X	X	X
Creation of preservation metadata	X	X	X
Secure storage and backup	X	X	X
Regular fixity checks	X	X	X
Regular virus checks	X	X	X
Periodic refreshment to new storage media	X	X	X
Transformation to a more preservable format	N/A	X	
Storage of original digital object	X	X	X
Strategic monitoring of format for changes	X	X	
Migration to successive format upon obsolescence	X		

Edit | Attach | Print version | History:
r17 < r16 < r15 < r14 < r13 | Backlinks | Raw View | WYSIWYG edit | More topic actions

JOHNS HOPKINS UNIVERSITY
What is the Data Conservancy?
http://dataconservancy.org/

Data Conservancy

A Blueprint for Research Libraries

ABOUT PARTNERS OBJECTIVES NEWS AND EVENTS RESOURCES CONTACT

What is the Data Conservancy?

Complex interactions among the atmosphere, the ocean, the land, the biosphere, and human behavior pose daunting challenges in understanding the causes of observed phenomena such as climate change and its associated impact on biodiversity and urbanization. Through collection, preservation, and semantic integration of data that are now very difficult to assemble and analyze, the Data Conservancy will transform the ability of scientists to answer grand challenge questions that are important to the nation and the world.

The Data Conservancy will research, design, implement, deploy, and sustain data curation infrastructure for cross-disciplinary discovery with an emphasis on observational data. Initial efforts of the project will highlight astronomy, earth sciences, life sciences, and social sciences.

The Data Conservancy is sponsored by the National Science Foundation under DataNET award OCI0830976.

Visit us on: **Facebook** • **YouTube** • **Twitter**

THEME DESIGN Copyright © 2009 under GPL by Tribute Media. Site Powered by Drupal.

JOHNS HOPKINS UNIVERSITY
Data Conservancy: Objectives
http://dataconservancy.org/objectives

Data Conservancy
A Blueprint for Research Libraries

ABOUT PARTNERS OBJECTIVES NEWS AND EVENTS RESOURCES CONTACT

Home

Objectives

In 2009, Data Conservancy answered the National Science Foundations call to create a world where "digital data are routinely deposited in well-documented form, are regularly and easily consulted and analyzed by specialists and non-specialists alike, are openly accessible while suitably protected, and are reliably preserved."

In their charge to reach this goal, DC researchers have revealed that cross-disciplinary research cannot be accomplished by the imposition of technical standards but rather through careful negotiation, long-term understanding, and localized demonstrations of benefits for current scientific problems. This realization has prompted DC to acknowledge the size of the problem space and include proofs of concept as part of our prototyping.

Implementation

Data Conservancy (DC) utilizes a unique approach reflective of our overarching vision through four major components: diversity of domain sciences, data preservation, educational programs, and library-led organizational framework.

1) **Diversity of Domain Sciences –** By engaging a varied spectrum of science domains – astronomy, earth sciences, life sciences, and social sciences, DC is able to explore various types of data structures, deep investigations of the exemplar community of astronomy, and broad examinations of disciplines typically described as "small science." While this multi-disciplinary nature represents an important opportunity to comprehensively examine scientific needs, identifying the optimal pathway for supporting and promoting interdisciplinary science requires additional research.

2) **Data Preservation –** DC architecture maps onto the Open Archival Information System (OAIS) reference model for digital preservation and thus allows the DC infrastructure to ingest data through multiple modes.

Preservation entails preparing content with appropriate representation information, context, metadata, and fixity such that someone other than the producer of the data can access, use, and properly interpret them. It also ensures long-term preservation of data through not only the bits, but also format and media migrations, or other actions consistent with an overall policy framework. DC's approach should allow for remote curation of data, use of preservation services from other providers, and flexibility as research results inform technical future infrastructure development.

3) **Capacity Building –** DC emphasizes workforce development and broadening participation through a unique set of educational programs. DC features a comprehensive set of institutes, summits, fellowships, and internships related to education and outreach. DC partners Illinois and UCLA have developed new courses and curriculum around the paradigms of Data Conservancy. Additionally, DC utilizes data scientists as mentors for recent graduates, post-docs, and Library staff persons acting as data scientists. This cluster of data scientists embedded in multiple science teams and the Library represents a unique opportunity to learn lessons and apply findings from across the DC network.

4) **Libraries as Cornerstone for Sustainable Infrastructure –** *DC will represent a blueprint for research libraries in the data age.* DC will bolster the resource base and capacity of research libraries toward data curation. In cases where research libraries may not have local capacity, DC could provide services on a fee basis thereby providing another revenue stream.

The Sheridan Libraries (SL) at Johns Hopkins University has been a leader of digital libraries and preservation for over a decade and has already led a major long-term funding effort that has resulted in a draft sustainability plan, an initial service stack, and a business plan that has been submitted to JHU central administration.

Vision for the Next Three Years

The first 18 months of DC were focused on prototyping, which have created the foundation for full-fledged preservation, improved conduct of science, and developed greater insights into current science and frameworks for new forms of science. In the next three years, DC will:

- Augment the open and flexible architecture for data curation and data synthesis.
- Extend the current data model or define new data models.
- Develop additional pilots and proofs of concept.
- Research the full problem space of CI development and cross-disciplinary science.
- Strengthen connection points between DC socio-technical research and infrastructure.
- Create a DC operational environment that provides data management support.
- Build capacity through continued community engagement of various stakeholders.
- Expand upon initial sustainability planning through case studies and further market analysis.

The Data Conservancy is sponsored by the National Science Foundation under DataNET award OCI0830976.

Visit us on: **Facebook** • **YouTube** • **Twitter**

THEME DESIGN Copyright © 2009 under GPL by Tribute Media. Site Powered by Drupal.

Sustainability of Digital Formats
Planning for Library of Congress Collections

Go

The Digital Formats Web site provides information about digital content formats. The analyses and resources presented here will increase and be updated over time. The compilers, Caroline R. Arms, Carl Fleischhauer, and Jimi Jones invite feedback on the content.

Introduction

Background information and overview: What is a format? How shall we evaluate formats? What projects in other organizations are addressing these questions? >>

Overview | Formats, Evaluation Factors, and Relationships | Papers and Presentations | Related Resources

Sustainability Factors

What affects the ability of the Library to preserve content in a given format? These sustainability factors apply to all formats. >>

Disclosure | Adoption | Transparency | Self-documentation | External Dependencies | Impact of Patents | Technical Protection Mechanisms

Content Categories

The evaluation of formats must take into account quality and functionality. These factors vary according to the type of content under consideration and the categories will be expanded as time passes.

Note that descriptions also exist for Generic Formats.>>

Still Image | Sound | Textual | Moving Image | Web Archive | Datasets

Format Descriptions

Documents with more information about specific formats. >>

Browse categories | Browse alphabetical list

Last updated 01/31/2011

UNIVERSITY OF MICHIGAN

What is Digital Preservation?

http://www.lib.umich.edu/preservation-and-conservation/digital-preservation/what-digital-preservation

MLibrary

About MLibrary Services Libraries | MGet It Search Tools Catalog (Mirlyn)

UNIVERSITY OF MICHIGAN Search GO Browse Get Help Ask a Librarian Using the Library

Preservation & Conservation

- EMERGENCY RESPONSE
- Book Repair
- Bindery Preparations
- Conservation
- Digital Preservation
  - Digital Preservation Glossary
  - External Resources
  - Internal Resources
  - What is Digital Preservation?
- Environmental Monitoring
- Mass Deacidification
- ALEPH Status Codes
- Directions
- FAQs
- Handouts and Links
- Contact Us

Home > Services > Preservation and Conservation > What is Digital Preservation?

What is Digital Preservation?

Introduction to the Problem

Today, we have access to information and data that 15 years ago would have scarcely seemed possible. It seems that almost everything is being created and used in the digital realm. Documents such as your history report, the spreadsheet that shows last year's travel budget and more were likely all generated on your computer. However, though we use computers for so many things, we often don't give much thought about preserving what we do generate until it is too late. Most people can remember at least one horror story of lost data, whether it happened to them or to a friend; the research paper that was lost when the computer crashed or the scattered and disorganized family photos that were only saved to one hard drive – that eventually crashed! This list of lost digital data illustrates the potential fragility of digital information. There are several reasons why digital objects are so fragile.

Fragility of Digital Objects

One reason that digital information is fragile is that software programs and other technologies can be very quickly superseded by newer ones and fall out of use. This phenomenon is called technological obsolescence. Once newer technologies become accepted as the norm, it can be difficult to use any digital object that exists in an older format. Although there is currently some backwards capability available for popular programs – for example, Open Office is able to open a Microsoft Office 2003 doc – this is not necessarily the case for less widely used programs and proprietary formats from small companies. Obsolescence can also occur with the media that digital information is stored on. It is quite difficult now to find a computer with a 3 ½" floppy drive, much less one for 5 ¼" floppies. These obsolete media or formats may contain unique information that may be very difficult or impossible to recover.

Another problem associated with digital preservation is media degradation. The very media that digital information is stored on was not always made to last, and can quickly degrade. This media can include magnetic tapes, floppy discs, optical discs and more. Take, for example, a movie on DVD. More likely than not, you have experienced a crucial scene in a movie being ruined because of scratches on a DVD you were watching. This is a case of media degradation – the information that was on the DVD no longer exists because the media that it was on has itself degraded. Now imagine this has happened with not simply a commercially available CD or DVD, but a unique item that contained thousands of digitized images. Clearly there are many risks associated with media degradation, especially when you consider how much information that has been burned onto CDs and DVDs to serve as a backup.

What is Digital Preservation?

So how do we deal with the problems mentioned above? One way we can do this is through active Digital Preservation. Digital Preservation is the management and maintenance of digital objects (the files, or groups of files, that contain information in digital form) so they can be accessed and used by future users. It is important to start thinking about digital preservation early in the life cycle of a digital object because while traditional print objects may last relatively unharmed for decades untouched, this is not the case with digital objects, which have significantly shorter life spans. Therefore, by thinking about preserving the digital object early on, even when it is created, we save a great deal of time and stress later on when trying to retrieve the information an object holds before it is too late. In this sense, digital preservation, and especially early digital preservation, is important not only for personal data management but also large repositories that manage many objects. Though personal horror stories of lost data seem to be scattered and only happen from time to time, for larger repositories that contain many hundreds and thousands of digital objects, lost data can be a much bigger problem. Digital Preservation, after all, is frequently focused on long term use, which can be quite difficult to achieve considering how fragile digital objects can be.

There are several strategies used to help preserve digital objects, such as emulation, migration and data redundancy.

Digital Preservation Strategies

One of the best ways to help preserve digital objects is by data redundancy. This is, simply put, making sure there are many copies of important files. If there are one or more copies of an important file available, it mitigates the disaster of the computer crashing or one disc being lost. However, though this may be helpful in the short term, it may not prove to be helpful in the long term, as file formats and media can change rapidly over a short period of time. In this case, two more digital preservation strategies can be helpful in preserving digital objects, emulation and migration.

Emulation involves using a program that imitates the original, obsolete hardware or software to render a digital object. In emulation, the original bit stream (the information that comprises the file) is saved and used. In contrast, in migration, the original bit stream is changed over to a new, current file format. Both strategies allow for the use of digital objects that may require outdated software or hardware, but in slightly different ways. When choosing a strategy, it is important to consider how the digital objects are to be used as well as the significant properties of that object. For example, is it a word document where you only need to read the information contained in it? In this case, migration which would eliminate some of the formatting might be ok. But what about a computer game where migrating data instead of emulating it would cause significant changes to the way the game was played? Although there are merits to both strategies, these types of questions are good to ask before choosing one. A more in-depth comparison of these two strategies can be seen below.

Emulation	Migration
• Can retain 'look and feel' of original digital object • Focus is on recreating the experience, not just accessing the content • Preserve and use original digital object. • Emulator will also have to be preserved + will need to update periodically • Without original hardware/software, you can't be sure you're retaining the exact look & feel • Does not always result in a perfect presentation of original digital object	• Can retain 'look and feel' of original digital object, depending on migration strategy as well as the format being migrated to • May lose original formatting, causing the object to not look quite the same as it did • Focus is on making the content available • May or may not save original digital object for backup/future migration purposes. File made in migration is a new copy.

One last way to help preserve digital objects is to make sure that as much information as possible is gathered when they are created. This information is called metadata and can include basic descriptive information about the file as well as information about the file format of the object. The metadata collected about an object helps to place items in context, as well as give specific information. This is essential for making sure that digital objects are authentic. Authenticity is that the file hasn't been added to or modified in any way. This means that it is the digital object created by the producer and the content of the digital object was not modified once it was placed in the digital repository. This is especially important for digital files that can be easily changed in a way that may not be easily apparent as opposed to print media. In addition, metadata can also help to track what was done to preserve the object throughout its life cycle, such as migrating an object from one format to another. This metadata can be linked to the digital object or encapsulated with the digital object itself. Encapsulating the metadata with the object, for example placing the metadata with the object in the same folder in a zip file, ensures that the information stays with the file no matter where it goes. Linking the metadata and storing the metadata somewhere else (not with the file), ensures that the information about the file can be recovered even if the object itself was lost.

UNIVERSITY OF NORTH CAROLINA

Carolina Digital Repository. Preservation and Stewardship

https://cdr.lib.unc.edu/external?page=about.preservationAndStewardship

CAROLINADIGITALREPOSITORY

Search all collections Go

Home | Browse | About | Contact | Login | Advanced Search

Preservation and Stewardship

The Carolina Digital Repository (CDR) manages the infrastructure and services necessary to provide sustainable access to digital objects.

The CDR is designed and operated to ensure the integrity of digital files at a bitstream level (the way information is encoded). Characteristics that will affect preservation are recorded in metadata as part of the deposit process. The CDR regularly verifies the integrity of files, maintains a record of preservation-related actions, and employs best practices in the field for persistent storage, including back-up and recovery procedures.

Repository staff consult with prospective depositors regarding the nature of their digital collections, foreseeable challenges for long-term access, and strategies for meeting their preservation goals.

Services Provided

- Consult with pilot collection contributors for preparation of submission of content into the CDR
- Capture descriptive information about digital objects
- Create and maintain service application, including:
  - Search
  - Display
  - Export
- Provide and manage sufficient storage for ingest and maintenance of content and associated metadata
- Provide and manage persistent storage, including appropriate back-up and recovery procedures
- Sustain bitstream-level preservation of digital objects
- Perform system monitoring, testing and debugging

Caveats

- Services provided by the CDR reflect the repository's stage of development and implementation of evolving standards and best practices in the field. As the repository grows and the technology matures, so will the CDR's capabilities for preservation and access.

Home | Browse Collections | About | Contact Us | Library Home | UNC Home

UNIVERSITY OF UTAH

Digital Preservation. Program Overview

http://www.lib.utah.edu/collections/digital/digital-preservation.php

J. Willard Marriott Library
THE UNIVERSITY OF UTAH

A - Z | library catalog | floor maps | hours

Search

HOME | RESEARCH TOOLS | SERVICES | COLLECTIONS | ABOUT THE LIBRARY | GET HELP

Home » Collections » Digital Collections » Digital Preservation

Digital Preservation

Program Overview | Organizational Policy Framework | Appendices | Preservation Decision Checklist

Digital Preservation Program Overview

The J. Willard Marriott Library has a mandate to preserve its unique collections in whatever form they exist. When it comes to digital materials, this can be a challenge because they are inherently fragile and can be difficult to maintain in the present while making sure they stay readable and useful for future generations.

In response to these challenges, we are creating a Digital Preservation Program within the Library. The mission of the program is to preserve and sustain long-term accessibility to unique digital collections housed within the Library.

The first step towards the new program has been to create a digital preservation policy framework, which is intended to be the highest-level digital preservation document at the Library. The framework makes explicit the objectives and priorities of the program.

As the program evolves, we will update this page to reflect the continuing work we are doing in this area.

Contact the Library's Digital Preservation Archivist for more information.
tawnya.keller@utah.edu or 801.581.8594

Contact

Special Collections
801-581-8863
801-581-3886
Digital Technologies
Office: 801-585-5587

Related Links

Prices
Equipment & Facilities

J. WILLARD MARRIOTT LIBRARY, 295 S 1500 E SLC, UT 84112-0860 | 801.581.8558 • FAX 801.585.3464
© THE UNIVERSITY OF UTAH | CONTACT US | DISCLAIMER | PRIVACY | STAFF INTRANET | LAST MODIFIED 9/29/11

Find the Marriott Library on:

UNIVERSITY OF UTAH
Organizational Policy Framework
http://www.lib.utah.edu/collections/digital/digital-preservation.php

J. Willard Marriott Library
THE UNIVERSITY OF UTAH

A - Z | library catalog | floor maps | hours

Home Collections Digital Collections Digital Preservation

Digital Preservation

Program Overview | Organizational Policy Framework | Appendices | Preservation Decision Checklist

PURPOSE
The J. Willard Marriott Library (hereafter, Library), in keeping with its mission, serves as a trusted caretaker of the Library's collections of enduring value1, including those in digital format. The Digital Preservation Policy Framework supports this mission and is the highest level digital preservation policy document at the Library. The framework makes explicit the Library's commitment to preserving its digital collections through a comprehensive digital preservation program for both born- analog and born-digital collections. The framework reflects the goals defined in the Library's SMART goals 2008-2009 and contains references to other relevant Library policies and procedures. The audience for the framework includes Library employees, digital content contributors, donors, and users.

MANDATE
Although many programs and projects both within and outside the Library make objects available to users online, digital preservation implies more than making an object available in a digital format. Digital preservation has been defined by the American Library Association (ALA) as "policies, strategies, and actions to ensure access to reformatted and born digital content regardless of the challenges of media failure and technological change. The goal of digital preservation is the accurate rendering of authenticated content over time."

The mandate for digital preservation at the Library is linked to institutional responsibility, legal obligations, scholarly commitment, contractual obligations and grants, and membership services (such as Utah Academic Library Consortium (UALC), Greater Western Library Alliance, Mountain West Digital Library (MWDL), etc). Special Collections, Information Technology, University Archives and the Institutional Repository all have missions, whether explicit or implicit, to collect, preserve, and provide access to the historical collections and institutional and scholarly records they hold. In some cases analog preservation will not suffice and the digital preservation of such objects can be inferred.

Additionally, the Library receives grant funding to ensure that specific collections are digitized and made available to online users and the sustainability and long-term accessibility of those collections is often required.

The Library also provides services for outside institutions that need items digitized and made available online. As part of these services, the long–term preservation of selected materials has been written into many formal agreements.

View Complete Framework

Contact

Special Collections
801-581-8863
801-581-3886
Digital Technologies
Office: 801-585-5587

Related Links

Prices
Equipment & Facilities

J. WILLARD MARRIOTT LIBRARY, 295 S 1500 E SLC, UT 84112-0860 | 801.581.8558 • FAX 801.585.3464
© THE UNIVERSITY OF UTAH | CONTACT US | DISCLAIMER | PRIVACY | STAFF INTRANET | LAST MODIFIED 9/29/11

Find the Marriott Library on:

UNIVERSITY OF UTAH

J. Willard Marriott Library

Digital Preservation Program: Organizational Policy Framework

(Tawnya Mosier, June, 2010; Updated by Tawnya Mosier Keller and Lisa Chaufty, September, 2011)

SECTION A ..... 2 - 5
Purpose
Mandate
Objectives
Scope
Attributes and Responsibilities
Challenges and Incentives
Operating Principles
Roles and Responsibilities
Cooperation and Collaboration
Selection and Acquisition
Access and Use

SECTION B ..... 6 - 12
Policies and Procedures
Roles and Responsibilities
Collection Owner Responsibilities
Collection Manager Responsibilities
Digital Preservation Program Responsibilities
Quality Creation and Benchmarking
Selection and Acquisition Policies and Procedures
Transfer Requirements and Deposit Guidelines
SIP Requirements
Deposit Agreement Requirements and Responsibilities
Access and Use Policies
Digital Archive Operations
Platform Requirements and Procedures

SECTION C 13- 14
Implementing the Framework
Publicizing and Promulgating the Policy and Plan

APPENDICES .....15 - 24
A: Definitions
B: Digital Preservation Decision Chart
C: Supported Formats Table
D: De-accession Criteria
E: Current Security, Emergency Planning, Platform Requirements and Procedures
F: Detailed Functional Model of the OAIS Reference Model
G: References

SECTION A

PURPOSE
The J. Willard Marriott Library (hereafter, Library), in keeping with its mission, serves as a trusted caretaker of the Library's collections of enduring value[1], including those in digital format. The Digital Preservation Policy Framework supports this mission and is the highest level digital preservation policy document at the Library. The framework makes explicit the Library's commitment to preserving its digital collections through a comprehensive digital preservation program for both born-analog and born-digital collections. The framework reflects the goals defined in the Library's SMART goals and contains references to other relevant Library policies and procedures. The audience for the framework includes Library employees, digital content contributors, donors, and users.

MANDATE
Although many programs and projects both within and outside the Library make objects available to users online, digital preservation implies more than making an object available in a digital format. Digital preservation has been defined by the American Library Association (ALA) as "policies, strategies, and actions to ensure access to reformatted and born digital content regardless of the challenges of media failure and technological change. The goal of digital preservation is the accurate rendering of authenticated content over time."

The mandate for digital preservation at the Library is linked to institutional responsibility, legal obligations, scholarly commitment, contractual obligations and grants, and membership services (such as Utah Academic Library Consortium (UALC), Greater Western Library Alliance, Mountain West Digital Library (MWDL), etc). Special Collections, Information Technology, University Archives and the Institutional Repository all have missions, whether explicit or implicit, to collect, preserve, and provide access to the historical collections and institutional and scholarly records they hold. In some cases analog preservation will not suffice and the digital preservation of such objects can be inferred.

Additionally, the Library receives grant funding to ensure that specific collections are digitized and made available to online users and the sustainability and long-term accessibility of those collections is often required.

The Library also provides services for outside institutions that need items digitized and made available online. As part of these services, the long-term preservation of selected materials has been written into many formal agreements.

[1] The J. Willard Marriott Library has defined enduring value as unique materials concerning Utah life and/or history as well as materials created by University faculty or Utah residents that fit the Library's collection mission.

OBJECTIVES
The overall mission of the digital preservation program is to preserve and sustain long-term accessibility to all digital collections created or collected throughout the Library by maintaining a comprehensive digital preservation program. Additionally, it should be noted that in order to manage digital collections over time, the program must include the accessibility of the software and other discovery tools associated with those collections.

Within the overall mission, we have the following objectives:

- Enable uninterrupted (not necessarily instant) access to digital content over time as technology for digital content evolves.
- Collaborate with campus partners and regional and national institutions to make the best use of resources and avoid duplication of effort.
- Comply with and contribute to the development of the standards and best practices of the digital preservation community.

SCOPE
The Library has primary responsibility for preservation of:

- Digital library resources of enduring value
- Digital resources from outside sources that the Library has contracted to preserve for long-term access

Program limitations: This program's top priority will not be to preserve objects that are already commercially available elsewhere or that are preserved with a trusted digital repository, except in the case of a future digital preservation strategy (such as the LOCKSS model).[2] The program will assess candidates for digital preservation within budget limitations as well as explicit criteria specified by the Library's Digital Collections Policy and tool (url forthcoming).

Program priorities:

- Unique materials in danger of obsolescence in analog form and identified as "critical need" for digital preservation
- Unique materials in digital form in danger of obsolescence or loss.
- Digital collections earmarked by our patrons as requiring long-term access

Timeframe: Our policy, procedures, current and needed technical infrastructure, refined selection criteria, and resources framework will be completed in 2012. At that

[2] There may be cases in which the program will archive an object that is also available within another trusted repository in order to retain the integrity of the collection. An example of this would be if a faculty member deposits her research with the University of Utah's Institutional Repository (USpace) but also deposits or publishes some research elsewhere. Another example would include a case in which the Library digitizes a rare book from its collection and Google Books digitizes the same book at a later date. Although that content may be duplicated, the Library would have a responsibility to retain its copy of the digitally preserved book because Google Books has made no claim to be a Trusted Digital Repository and therefore its content is not guaranteed to be available in perpetuity.

point, we will assess the overall timeframe for an operational, sustainable, comprehensive digital preservation program.

ATTRIBUTES and RESPONSIBILITIES
This framework follows digital preservation standards as defined in OCLC's Trusted Digital Repositories: Attributes and Responsibilities. Accordingly, the attributes of a trusted digital repository are:

- Open Archival Information System (OAIS) compliance
- Administrative responsibility
  - Accept responsibility for the long-term maintenance of digital resources on behalf of its depositors and for the benefit of current and future users.
- Organizational viability
  - Establish an organizational system that supports not only long-term viability of the repository, but also the digital information for which it has responsibility.
- Financial sustainability
  - Demonstrate fiscal responsibility and sustainability.
- Technological and procedural suitability
  - Develop policies, practices, and performance that can be audited and measured.
- Systems security
  - Ensure the ongoing management, access, and security of materials deposited within it.
- Procedural accountability
  - Dependably carry out its long-term responsibilities to depositors and users openly and explicitly.

CHALLENGES and INCENTIVES

- Budget limitations. We must always live within our financial means. Realistically, we will not be able to preserve everything, making our selection criteria for preservation all the more imperative.
- Keeping up with technological change in terms of hardware, software, new formats, etc. A key question here deals with emulation vs. migration of formats.
- Creating and following submission standards
- Meeting the education needs of staff involved *with* (but not explicitly responsible *for*) digital preservation.

OPERATING PRINCIPLES
The Library will strive to:

- Comply with OAIS and other digital preservation standards and practices
- Ensure that content remains readable and understandable
- Participate in the development and adoption of digital preservation community standards, practice and solutions

- Develop a reliable, scalable, sustainable, and auditable digital preservation repository
- Manage the hardware, software, and storage media in accordance with environmental standards, quality control specifications, and security requirements

ROLES and RESPONSIBILITIES
The Library accepts responsibility for preserving its digital assets. The Technology Services Council evaluates high-level policy documents and reviews programmatic plans and progress. The Associate Director for Information Technology and the Associate Director for Scholarly Resources and Collections provide input and guidance to the work being done by the Digital Preservation Archivist to manage the digital preservation program and the lifecycle of digital objects of enduring value within the Library. The Head of Digital Ventures, Head of University Archives and Records Management, and the Institutional Repository Coordinator also contribute to the program at various levels.

COOPERATION and COLLABORATION
The Library acknowledges that its digital preservation goals will likely exceed available resources and therefore not be able to guarantee the safety of all digital assets. Therefore, collaboration and partnerships with regional and/or like-minded organizations will be required to ensure the program's success and to properly prioritize which assets will be addressed and in what order. These may include working with state and regional cultural heritage organizations. Such collaborations may require formal agreements that make explicit the roles and responsibilities of each member in any collaborative.

SELECTION and ACQUISITION for PRESERVATION
The Digital Preservation Decision Chart (Appendix B) guides collection owners regarding preserving digital content of enduring value. The Decision Chart also reflects criteria for deposit.

ACCESS and USE
Stakeholders of the Library's digital preservation program include traditional users such as Library departments, patrons, and faculty, and newer stakeholders such as the University and cultural heritage organizations that have deposited archival masters with the Library for long-term preservation. Restrictions to use of collections are defined by the collection holder and vary from collection to collection.

Archive Links

Go to the Archive
Mediated Submission Form

About the Archive

What is the Archive?
FAQ

Policies

Access Restriction Policy
Digital Collection Policy
Copyright and Author Rights Policy
Digital Preservation Policy
Withdrawal Policy

Questions?

Contact us at:
rworks@uw.edu

ResearchWorks Archive Digital Preservation Policy

What is digital preservation?

Digital preservation is the series of managed activities necessary to extend the usable life of computer files and protect them from media failure, physical loss, and obsolescence. Digital preservation can be divided into activities that maintain the bitstream or file and activities that maintain access and functionality essential to the purposes for which the original digital material was created or acquired (allowing the file to be opened and viewed with the same look and feel as the original creator intended). At this time, ResearchWorks Archive is committed to preserving the bitstream of the digital object. Further measures to maintain access and functionality over time will be taken as resources permit.

What ResearchWorks Archive does to support digital preservation

To preserve and provide access to the bitstream, ResearchWorks Archive:

- Maintains multiple copies of each digital object across multiple sites
- Maintains the authenticity of the bitstream through integrity checking
- Maintains data security according to industry standards
- Monitors storage media and copies data to new storage media as needed
- Provides sufficient metadata and persistent identifiers to provide reliable access to digital objects
- Supports multiple file formats but the level of preservation support will vary depending on the file format

Levels of Preservation Support

The level of preservation support will depend on the file format, the application used to generate the content, and the set of features used. While efforts will be made to preserve work in any digital format, submission in a recommended file format is strongly encouraged.

File formats with the following characteristics will more likely be able to retain their functionality over time:

- Complete and open documentation
- Platform independence
- Wide adoption
- Non-proprietary (vendor-independent)
- No “lossy” or proprietary compression
- No embedded files, programs or scripts
- No full or partial encryption or password protection

In addition to the ResearchWorks Archive List of Preferred File Formats, guidance for creating “preservation-friendly” digital resources is available from the following sources:

- JISC Digital Media: http://www.jiscdigitalmedia.ac.uk/
- Library of Congress Sustainability of Digital Formats: http://www.digitalpreservation.gov/formats/
- Federal Agencies Digitization Guidelines Initiative: http://www.digitizationguidelines.gov/

Digital preservation strategies are still evolving. To insure discovery and interoperability, the UW Libraries will follow standards and best practices where they exist, from creation throughout the lifecycle of the digital resource.

An effective digital preservation program will ensure continued access to digital materials for as long as necessary.Preservation decisions will be made within the context of the Collection Policy, balancing scholarly and historical value, user accessibility, and cost constraints.