


8-23-2017

Policy - Scholars' Mine Digital Preservation

James Roger Weaver

Missouri University of Science and Technology, weaverjr@mst.edu

Follow this and additional works at: http://scholarsmine.mst.edu/scpro_guidelines

 Part of the [Archival Science Commons](#), and the [Scholarly Communication Commons](#)

Recommended Citation

Weaver, James Roger, "Policy - Scholars' Mine Digital Preservation" (2017). *Scholars' Mine - Policies, Procedures and Guidelines*. 5.
http://scholarsmine.mst.edu/scpro_guidelines/5



This work is licensed under a [Creative Commons Attribution-Noncommercial 4.0 License](#)

This Documentation is brought to you for free and open access by Scholars' Mine. It has been accepted for inclusion in Scholars' Mine - Policies, Procedures and Guidelines by an authorized administrator of Scholars' Mine. This work is protected by U. S. Copyright Law. Unauthorized use including reproduction for redistribution requires the permission of the copyright holder. For more information, please contact scholarsmine@mst.edu.

Policy – Scholars' Mine Digital Preservation

Purpose

The Library is committed to preserving and maintaining digital resources that support the mission of the University and Library for as long as they are useful to their respective community of users. This policy describes the need and strategies for preserving Missouri S&T's digital resources and outlines the proactive steps taken to preserve these materials to ensure that faculty, staff, students, and other users will have ongoing sustainable access to our expanding digital collections. Collection specific strategies for continuous preservation will be developed in conjunction with this policy. This policy is subject to change as capabilities, standards, best practices and technology impact the Library's ability to preserve these resources. The intended audience for this policy includes:

- faculty, administrators, and students of Missouri S&T,
- researchers external to Missouri S&T who are collaborating with Missouri S&T personnel and who want to make use of Scholars' Mine services,
- funders, and
- users.

The objectives are:

- to collect, publish and preserve the digital resources created by our campus community as they relate to research, education, and campus culture and life. This includes the digital data sets and associated documentation generated by campus community researchers and their affiliates.
- to enable campus community researchers to satisfy the requirements of funding agencies in managing, sharing and preserving research data.
- to provide the means for researchers, policy makers, and others to discover and access digital resources, including datasets, generated through research done at or in conjunction with Missouri S&T for the long term.
- to provide a sustainable preservation environment where digital resources including research datasets, are available to support the historical record of research and are accessible for use for contemporary scholarship.

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

- Describe the challenges associated with digital preservation.
- Explain the necessity of a digital preservation policy.
- Outline the principles on which digital preservation actions are based.
- State the scope of digital preservation activities, including sources and types of digital content that will be preserved.
- Define the strategies that will be performed to ensure the long-term preservation of digital resources.
- Identify stakeholders responsible for components of the digital preservation strategies.

- Define a policy review schedule.

Introduction

Curtis Laws Wilson Library creates and manages original digital content as well as purchases and licenses a growing number of digital resources. Digital objects have a much higher preservation risk compared to traditional materials due to their fragile nature and evolving hardware, software, standards, and file formats. The library defines digital preservation as "the series of management policies and activities necessary to ensure the enduring usability, discoverability, and accessibility of content over the very long term." Digital preservation differs from analog preservation in that:

- Digital preservation requires active management. While many analog materials, such as books, can survive for years when simply stored in a climate-controlled environment, digital materials that are left alone for long periods of time are much more likely to degrade beyond repair, and this degradation is generally not discovered until there is an attempt to use the item.
- The preservation needs of analog materials, such as books, journals, film, and tape, are well understood and have not greatly changed over time. However, digital preservation is a new and developing field with standards that are still being created. New tools and technologies will require that digital preservation activities be responsive and adaptable.
- The expertise to treat analog materials generally exists within one department within the library. The expertise and actions required to preserve digital content exists across multiple library departments as well as external to the library. A robust digital preservation infrastructure will inherently operate within a collaborative and communicative workspace

Mandate

As an institution of higher learning Missouri S&T supports scholarship, teaching, and learning. Curtis Laws Wilson Library is responsible for identifying, acquiring, and preserving scholarly resources for as long as they are needed to support the activities of the University. The digital repository's (Scholars' Mine) mandate for the preservation of digital resources is drawn from responsibilities outlined above and specifically defined in:

- Lever 2.4.13 of Missouri S&T's strategic plan which states: "Continue development of Scholars' Mine designed for increasing internal and external communication and globally promoting the intellectual work of the campus."
- Lever 5 of the library strategic plan which states: "Explore opportunities to feature and develop Scholars Mine."

Principles

These principles guide all digital preservation actions:

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

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

Collaboration: The Library will investigate and participate in collaborative agreements whenever possible so long as they are a good use of Library resources.

University and Library Missions: This policy and actions associated with it exist in support of stated University and Library missions. This policy will be annually reviewed against University and Library missions and goals to ensure that it continues to support the core work of the institution.

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

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

Sustainability: Digital preservation activities will be planned and implemented in ways that best manage current resources and can be sustained into the future.

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

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

Scope

Curtis Laws Wilson Library is responsible for identifying, securing, and providing the means to preserve and ensure ongoing access to digital resources. Digital resources collected by the library fall into these general categories:

1. Subscription-based resources: journals and databases, to which the Library pays for access, but does not own copyright.
2. Escrow files: archival backup files of electronic resources, which may be purchased by the Library or submitted by vendors in fulfillment of contractual obligations.
3. Missouri S&T-owned digital resources:
 - Analog objects owned by the Library or Missouri S&T that are selected for digital conversion.
 - Born-digital objects and publications created by the Library or the Missouri S&T community.
 - Electronic publications created by the Library or members of the Missouri S&T community as works for hire and with the assistance of Library staff and hosted by the Library.
 - Digital files that are produced in the course of creating a physical or digital exhibit by the Library or members of the Missouri S&T community.
 - Other Digital resources collected by the Library or Missouri S&T that are unlikely to exist elsewhere.
4. Selected research and scholarship generated by the Missouri S&T community:
 - Publications of faculty, students, research centers, labs and other community members.
 - Research data created by faculty, students, research centers, labs and other community members.
 - Other resources related to the research and scholarly activities of the Missouri S&T community.
5. Selected resources related to campus culture and life.

6. University records: records created by colleges, departments, and other units within the university in the course of conducting business.

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

Content Types

Each content source listed above may present content in one or more of the following content types. Each content type may require different preservation strategies.

Textural materials (ebooks, articles, etc.)

Images (scanned text, photographs, digital photographs, digital art, etc.)

Audio/video materials (videos, recorded sound, etc.)

Numerical data/datasets (research data, etc.)

The library may acquire materials in additional formats in the future which will require the development of additional preservation strategies.

Digital Preservation Strategies

The specific preservation strategy used for digital resources will depend on the source of and type of content, as well as existing technology, expertise, and ongoing support. Preservation actions to be taken are as follows:

Subscription-based resources: These resources are not owned or directly controlled by the Library; therefore Library cannot manage them. Instead, these resources are primarily managed by agreement with the publisher or vendor. The library will negotiate such preservation agreements when developing subscription and license contracts with publishers and vendors.

Resources created by, or for, and owned by the Library or University: These resources will be managed by the Library using the life cycle model outlined below. The expectation is that these resources and associated metadata will be developed according to current standards and best practices, and stored in a long-term repository within the Library infrastructure.

Scholarly and research resources created by members of our campus community, but not owned by the Library or University: These resources will be managed by the Library using the life cycle model outlined below. The expectation is that these resources and associated metadata will be developed according to current standards and best practices, and stored in a long-term repository within the Library infrastructure.

Life Cycle Management

Digital objects will be managed using the life cycle model. This model is a framework describing the stages that digital resources go through during their existence. The preservation of digital objects requires planning and action at every stage of an object's lifecycle.

Creation – As digital content is created, preservation actions should include creating and/or capturing administrative, descriptive, structural and technical metadata about the objects, as well as imposing a

well-defined storage system. Content will be created following current standards and best practices for capture and formatting.

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

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

Metadata Creation – All digital resources will adhere to the Library metadata policy and practice. Essential preservation metadata includes:

- Administrative
- Technical
- Structural
- Provenance
- Rights

Storage – Digital resources must be stored in a manner that is consistent with accepted best practices in the digital preservation community. This will include both technical infrastructure (hardware, software, network access, data backup, facilities, maintenance, etc.) and ongoing preservation management activities. Best practice in digital preservation requires duplicating digital objects in both local systems and geographically removed systems. The Library will pursue this by working with University IT and other service providers to host redundant local storage. Library staff will also explore other methods of storing data off site, such as in a private LOCKSS network, the HathiTrust, the Internet Archive, or another collaborative group.

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

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

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

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

may be performed include creating new content or metadata, adding content or metadata, migrating content to a new format, and creating a subset of content or metadata.

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

Levels of Preservation

The Repository provides three levels of preservation support for specific file formats. These support levels were determined by evaluating the prevalence of the file format in the marketplace, whether the format is proprietary, the availability of tools for emulation or migration and the availability of local resources to take specific preservation actions. The Repository will undertake appropriate format monitoring and provide adequate staffing and other resources to support the services offered at each level. Over time, we expect our ability to provide full preservation support for more formats to increase as additional tools and techniques are developed. We are developing specification and format best-practice guidelines for common content types in order to assist content providers and to ensure that contents meets the level of quality necessary for full information capture and the highest degree of preservability.

Levels of Preservation Support	
<p>Level 1 Best effort to maintain the content, structure and functionality in the future</p> <p>Example: TIFF</p>	<ol style="list-style-type: none"> 1. Provided only for formats that are publicly documented and widely used which makes it more likely that tools will exist or be developed to undertake preservation actions, and that such actions will result in understood and controlled transformation or migration. 2. Content may be transformed to another stable format (normalized) to provide assurance that the information content is preserved. 3. Content will be preserved as originally deposited to ensure the original bit stream is always available.
<p>Level 2 Limited efforts to maintain the usability of the files as well as preserving them as submitted</p> <p>Example: Microsoft Word</p>	<ol style="list-style-type: none"> 1. The file format will be monitored and may be transformed when significant risk to access is imminent. It may be difficult to predict or control the consequences of any transformation or migration on content, structure or functionality. 2. The file may also be transformed to a more preservable format to ensure that the information content is not lost, even if some structure and functionality are sacrificed. 3. This level of support is generally applied to proprietary formats that are widely used, where there is substantial commercial interest in maintaining access to files saved in the format, and therefore tools will likely be available to migrate them to successor formats
<p>Level 3 Basic preservation of the file and associated metadata as is</p> <p>Example: PhotoCD</p>	<ol style="list-style-type: none"> 1. No active effort will be made to monitor the file format, associate risks. or, to normalize, transform, or migrate the file to another format.

	<ol style="list-style-type: none"> 2. Files may be openable and/or readable by future applications, but there is no guarantee that the content, structure, or functionality will be preserved. 3. This service level usually applies to files written in highly specialized, proprietary formats, often usable only in a single software environment, formats no longer widely utilized, and/or formats about which little information is publicly available. 4. Any format not yet reviewed and evaluated will also receive Level 3 service on deposit. A higher level may be assigned after format review takes place.
--	---

Primary Preservation Services by Level			
Feature	Level 1	Level 2	Level 2
Persistent identifier that will always point to the object and/or its metadata	Yes	Yes	Yes
Provenance records and other preservation metadata to support accessibility and management over time	Yes	Yes	Yes
Secure storage and backup	Yes	Yes	Yes
Periodic refreshment to new storage media	Yes	Yes	Yes
Fixity checks using proven checksum methods	Yes	Yes	Yes
Storage in a trusted preservable format (making a normalized version, if necessary)	Yes	For Some Formats	No
Strategic monitoring of format	Yes	Yes	No
Migration to succeeding format upon obsolescence	Yes	No	No

The three levels of preservation are applied at the individual file level. Complex content items comprised of multiple files in various formats will require additional evaluation to determine whether the operational relationships between the files can be maintained. The original relationships can be documented externally in metadata to ensure that the relationships are preserved. Also executables and some files that rely on a specific hardware/software environment will require additional evaluation because not only the format but the access environment must be considered in making a preservation determination.

Stakeholders

Stakeholders in digital preservation include the Library, users of Library collections, faculty, students and other University community units and members who create digital content housed by the Library. Explicit responsibilities of stakeholders in carrying out preservation strategies include:

Acquisitions – Manages the purchasing and licensing of electronic resources.

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

Electronic Services Librarian – Manages the licensing of subscription-based digital content. Ensures that sufficient third-party preservation agreements are met whenever possible.

Collection Development Team – Manages the collection development review and de-selection of digital resources as needed. Ensures ongoing harmony of digital collections with print collections and the Library's collection development policies.

Archivist – (1) Manages University records, including ingestion of records into the records management system and subsequent transfer to Archives or other storage as needed. (2) Selects and manages College records and cultural artifacts to be preserved.

Scholarly Communications Librarian – (1) In cooperation with the Archivist manages the creation of digital content within the Library and ensures that standards and best practices are followed for the creation of digital content, including the capture of preservation metadata. (2) Oversees and manages the Library's digital preservation strategies, with particular emphasis on selection, ingestion, storage, preservation management, transformation, and coordination with third-party preservation services. Ensures general compliance with standards and best practices. Coordinates activities across departments and with external vendors. (3) Manages the Library's Digital Publishing Program.

Scholars' Mine Team – (1) Manages the creation, ingestion, and maintenance of digital objects into the Institutional Repository. (2) In coordination with other Library departments and Campus IT, manages the technical infrastructure needed to create, ingest, store, transform, and provide access to digital resources within the Institutional Repository. (3) Creates, installs, and maintains software as needed and provides support for staff using these tools. (4) Manages accessibility and user interface design to ensure usability and discoverability of digital resources.

Policy Review

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

Standards and Best Practices

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

Relevant standards include:

[Open Archival Information System Reference Model \(OAIS\)](#)

[PREMIS Data Dictionary for Preservation Metadata](#)

[Data Seal of Approval](#)

[Trustworthy Repositories Audit & Certification \(TRAC\)](#)