


23 Aug 2017

Policy - Scholars' Mine Research Data Management

James Roger Weaver

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

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

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

Recommended Citation

Weaver, James Roger, "Policy - Scholars' Mine Research Data Management" (2017). *Scholars' Mine - Policies, Procedures and Guidelines*. 6.

https://scholarsmine.mst.edu/scpro_guidelines/6



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 Research Data Management

Purpose

The Library is committed to preserving and maintaining research data created by our campus community as long as it supports the mission of the University and Library and is useful to the respective community of users. This policy supplements our existing Digital Preservation Policy by describing needs and strategies specific to preserving research data. Collection specific strategies for the continuous preservation of research data 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 research data created by our campus community that relates 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 research data generated through research done at or in conjunction with Missouri S&T for the long term.
- to provide a sustainable preservation environment where research data is available to support the historical record of research, and is accessible for use for contemporary scholarship.

This policy is a set of guidelines for the digital preservation of research data, 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 the digital preservation of research data.
- Explain the necessity of a research data management policy.
- Outline the principles on which the digital preservation of research data are based.
- State the scope of digital research data 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 research data.
- Identify stakeholders responsible for components of the digital preservation strategies.
- Define a policy review schedule.
- Define terms, identify standards, and list resources that will inform digital preservation activities.

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 such as research data 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 research data for as long as needed to support the activities of the University. The institutional repository's (Scholars' Mine) mandate for the preservation of research data 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 and Practices

These principles guide all digital preservation of research data:

Access: Digital preservation of research data 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 research data 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 research data 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.

Curtis Laws Wilson Library adheres to the following additional practices:

1. Seeks to comply with the Open Archival Information System (OAIS) reference model standard and other digital preservation standards and practices as they evolve.
2. Seeks to conform to certification requirements for ISO Standard 16363 for Trusted Digital Repositories.
3. Adheres to prevailing community standards for preserving access to digital content of long-term value so that it remains readable, meaningful, and understandable.
4. Commits to an interoperable, reliable, and scalable digital archive with appropriate storage management for content.
5. Documents policies, procedures, and practices clearly and consistently.
6. Maintains or contracts hardware, software, and storage media containing archival content in keeping with prevailing best practices, including adherence to environmental standards, quality control specifications, and security requirements.
7. Establishes adequate and secure backup and disaster recovery safeguards.
8. Establishes procedures to meet archival requirements pertaining to provenance, chain of custody, authenticity, and integrity of content.
9. Captures and maintains adequate metadata to document digital content and what is required to provide access to the content.
10. Defines a sustainability plan that ensures the cost-effective, transparent, and auditable management of the digital archive over time.
11. Seeks to monitor threats to the accessibility of digital content.
12. Complies with intellectual property, copyright, and ownership rights for preservation of and access to all content.
13. Attempts to allocate adequate and appropriate resources and infrastructure for sustained digital preservation, acknowledging the need to tailor collecting programs according to the ability to preserve content and the availability of resources for preservation activities.

Scope

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

1. Missouri S&T-owned research data:

Curtis Laws Wilson Library
Missouri University of Science and Technology

- Analog research data objects owned by the Library or Missouri S&T that are selected for digital conversion.
 - Born-digital research data objects created by the Library or the Missouri S&T community.
 - Other digital research data objects collected by the Library or Missouri S&T that are unlikely to exist elsewhere.
2. Selected research data generated by the Missouri S&T community:
 - Research data objects created in association to publications of faculty, students, research centers, labs and other community members.
 - Research data created by faculty, students, research centers, labs and other community members that exist independent of publications.
 - Other research data related to the research and scholarly activities of the Missouri S&T community.
 3. Selected research data related to campus culture and life.
 4. Research data related to University record: 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 research data throughout the research lifecycle 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.

Research data will be preserved and maintained for a minimum period of ten years. Longer-term preservation of research data beyond the ten year retention period is subject to:

1. requests for longer preservation periods by the researcher,
2. the Libraries' selection criteria and archival appraisal for long-term retention, and
3. budget approval for staffing and related resources needed to accomplish this goal.

Research Data Preservation Priority

Curtis Laws Wilson Library is committed to the preserving and providing access to the following classes of research data with associated preservation priorities and commitment levels:

Priority 1: Data Sets associated with Publications. Extensive effort will be made to ensure preservation of data sets associated with journal publications or other scholarly publications in perpetuity, or for as long as the data sets meet the Libraries collection development policies and practices, or is superseded in the future by an acceptable data repository.

Priority 2a: Stand-Alone Data Publications. Reasonable step will be taken to preserve stand-alone data publications in accordance with best practices and collection development policies.

Priority 2b: Data Sets with High Research/Teaching Value. Reasonable step will be taken to preserve data sets that are identified by faculty, subject specialist librarians or archivists as having high value for meeting the research and teaching needs of Missouri S&T or within the broader research community.

Priority 3: Other Data Files and Materials. No preservation steps will be taken for ephemeral materials deemed to be of little or no long-term value to the comprehensiveness of the collection. Working files of particular significance to Missouri S&T's teaching and research needs, or within the greater research community, may be preserved on a select basis as appropriate.

Content Types and File Sizes

There are no restrictions to the content types or file sizes for research data. However, there may be practical restrictions imposed by the technology and software employed by users accessing content.

Digital Preservation Strategies

The specific preservation strategy used for research data will depend on the source of and type of content, as well as existing technology, expertise, and ongoing support. Research Data will be managed by the Library using the life cycle model outlined below. The expectation is that the research data 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

Research data will be managed using the life cycle model. This model is a framework describing the stages that research data go through during their existence. The preservation of research data requires planning and action at every stage of the lifecycle.

Creation – As research data is created preservation actions should include creating and/or capturing administrative, descriptive, structural and technical metadata about the data, 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 of research data for 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 data selected for the library collections are intended for permanent retention unless explicitly stated otherwise.

Ingest – Ingestion of research data 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 research data will adhere to the Library metadata policy and practice. Essential preservation metadata includes:

- Administrative
- Technical
- Structural
- Provenance
- Rights

Storage – Research data 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, another data archive, or another collaborative group.

Preservation Management – A series of actions that will need to be performed on research data prior to and during long-term storage, and at varying levels depending on the source and type of data. 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 – Research data will be reviewed and managed to ensure that files are accessible into the future and the data is discoverable and created in a way that they may be easily found by all stakeholders.

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

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

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 research data 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](#)

[Digital Curation Center](#)

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