U-M Library Digital Repository Services Digital Preservation Policy

Last Updated: July 9, 2020

Introduction

The Dark Blue, Deep Blue, Digital Collections, and Fulcrum digital repository services comprise the Digital Repository Services Steering Committee (DRSSC) of the U-M Library. The DRSSC has authored this policy to cover content locally hosted in the digital repository services we manage.

Durability is among our core values, and the foundation of that durability lies in our commitment to the preservation of the content we provide access to. Our goal is to make access through our digital repository services synonymous with adding it to the durable record of scholarship; there

is no preservation without access, and there is no access without preservation. This policy expresses that commitment by answering the questions: "What can a future researcher or scholar expect to encounter in one of U-M Library's digital repository services? How do we faithfully represent today's content in the future?"

In the interest of transparency and trustworthiness, this policy summarizes the decisions and commitments Dark Blue, Deep Blue, Digital Collections, and Fulcrum digital repository services have made and anticipates the decisions we will have to make in the future. When concerns or questions regarding digital preservation occur, we will use this document to review the issue and take action. The DRSSC will review this policy every 2 years.

Definitions

- Designated community: An identified group of potential users who should be able to understand a particular set of information (OAIS, Section 1-11)
- Digital preservation: A combination of policies, strategies and actions to ensure access to digital content by addressing the challenges of media failure and technological change. The goal of digital preservation is the accurate rendering of authenticated content over time (adapted from ALCTS Definitions of Digital Preservation)
- Format: a file or bitstream's structure/"encoding"; can be expressed as a MIME type
- Retention review: a process by which works in repositories are evaluated to determine if the work will continue to be hosted in the repository and if so what preservation actions should be taken.
- Sensitive data: Data that typically cannot be made publicly available (see U-M's Sensitive Data Guide for a partial list), and may not be appropriate for preservation in any of our repositories. Note that determination of whether data are sensitive often requires judgment and discussion; potentially incendiary materials fall into this latter category
- Stakeholder: a depositor or creator of content to be preserved in one of our repositories
- Type: the type of content (e.g. article, audio, dataset, image) to be preserved
- Version of Record: the 'Version of Record' concept is intentionally malleable and subject to change as technologies and publishing conventions evolve, but is intended to capture the material that is essential to conveying the scholarly argument embodied in a work
- Work: a combination of one or more files plus metadata comprising an addition to one of the repositories; can also be expressed as a "Digital Object"

Digital Repository Service Descriptions

The descriptions below describe the roles and responsibilities of our current repository services, largely in terms of content.

Dark Blue

Dark Blue provides preservation storage in cases where dark storage is needed as part of an overall long-term access strategy. This includes in-process material not yet ready for access–such as born-digital material received through the Special Collections Research Center–and long-term preservation files–such as moving image preservation files.

Dark Blue deposits are restricted to defined content types deposited by authorized staff. Each content type has specific requirements for ingest and long-term preservation. Development of additional content types is ongoing and is managed by the Dark Blue Steering Committee.

Deep Blue Data

Deep Blue Data accepts data sets that (a) are produced by U-M faculty, staff, graduate students, or U-M academic and administrative research units with which our scholars are affiliated, (b) have research value and is not primarily administrative in nature, (c) do not include sensitive information or be subject to export controls, (d) are not encrypted.

Deep Blue Data allows self-deposit for small data sets; otherwise it is a mediated service. All deposits are reviewed by Deep Blue staff — both upon ingest and periodically afterwards — to assure quality and usability.

Deep Blue Documents

Deep Blue Documents accepts new and existing digital works identified as having current and lasting value. This work must be (a) produced or sponsored by U-M faculty, staff, students, or U-M academic and administrative research units with which our scholars are affiliated, (b) be educational, artistic, or research-oriented, and (c) a completed version, ready for distribution. The decisions on what the work should contain and offer are typically made by members of the U-M community at large.

Deep Blue Documents allows self-deposit of the materials described above, and provides batch loading services for larger deposits. Batch-loaded deposits are reviewed for metadata quality, copyright and file format compatibility; self-deposits typically are not mediated.

Digital Collections

Digital Collections accepts physically owned collections that are already digitized or in the process of being digitized, and makes them accessible through a unique, robust discovery interface (see all digital collections). These collections come from the U-M Library, Bentley Historical Library, William L. Clements Library, LSA Museums (specimen images), and the community. All materials are approved for hosting by the AUL for Collections at the U-M Library.

Digital Collections fully mediates the process of accepting, processing, and launching the materials in the digital collections. Staff work directly with a point person, usually a physical content curator.

Fulcrum

Fulcrum accepts the "Version of Record" of published works that are, in whole or in part, publications of Michigan Publishing imprints (University of Michigan Press, Michigan Publishing Services, ACLS Humanities E-Book) or contracted publisher partners (e.g. Amherst College Press, Lever Press, Against the Grain Media, British Archeological Reports).

Fulcrum fully mediates the process of accepting, processing and launching materials. At present, Fulcrum focuses on preserving the text of ebooks, any components embedded in the text (such as figures, tables, or time-based media), metadata (for the work as a whole, for the individual components, and relationships among the parts), and any supplemental materials that may be hosted on the platform.

Scope of our Preservation Commitment

Preservation always entails making choices about what to preserve. While we strive to capture the most comprehensive record possible of a hosted work, finite resources of time, attention, and technology require the scrupulous application of those resources to achieve a preservation strategy that is most likely to succeed technologically while being comprehensible and useful to future generations.

General Principles

- We accept and preserve content suited to the mission of the U-M Library – to support, enhance, and collaborate in the instructional, research, and service activities of the faculty, students, and staff, and contribute to the common good by collecting, organizing, preserving, communicating, sharing, and creating the record of human knowledge.
- We follow the principle of "Preservation is Access, Access is Preservation".
- We focus on preserving the content itself, believing that the user experience, including look and feel, can and should be improved with time.
- We preserve code as documentation of what was done, but not as something executable in current or legacy environments.
- We use metadata schemas according to library best practices, which include adhering to standards, using common controlled vocabularies, and enabling automated crosswalks wherever possible.

- Different types of content require different types of repository services. We create and maintain a variety of services because we recognize that content creators and end users have different expectations and needs.
- Each repository service has different guidelines and best practices for the content they preserve. See the service descriptions above for more on each service.
- We offer three levels of preservation via [U-M Library Digital Repository Services Registered Formats and Support Levels](#):
  - Level 1 is for select open and publicly documented formats we can **fully migrate** to new formats.
  - Level 2 applies to commonly used file formats we will **attempt to migrate**.
  - Level 3 applies to all files (including but not limited to the above), and we will provide **bit-level preservation**, maintaining the file exactly as-is when it was deposited into one of our services.
- We do bit-level preservation and migration to the best of our abilities.
- We reserve the right to remove, or not accept, content if we determine it's out of scope for our digital repository services (see: "What We Do Not Preserve" below).

What We Preserve

While each digital repository service sets its own policy for what content it accepts and retains, all content is subject to the same preservation policies. Our focus is on the files themselves and the levels of preservation we offer based on the file formats.

- There is almost always a home for something at the U-M Library. Some repository services set format requirements and others do not. We will direct stakeholders to the appropriate repository service, accordingly.
- When ingested, content should be complete and must be ready for distribution and re-use by others through the inclusion of appropriate descriptive metadata (and documentation as necessary).
- Sometimes we receive content that are in a variety of formats and types — we may decide to put different materials in different repositories.

It is important to note that our file format considerations apply to the preservation of all material deposited into the repositories. This includes supplemental material such as code books, "read me" documentation, etc.

The File Formats Working Group will be developing a monitoring service for file formats that will work in concert with the [U-M Library Digital Repository Services Registered Formats and Support Levels](#).

What We Don't Preserve

Though we make every effort to find a home for all content presented to us, we do have to turn away some material because of what it contains. If the material could place the creator, U-M Library, or the University of Michigan at risk we may need to reject its deposit at the outset.

Such concerns may also arise during a retention review, which we may do for any content in the repositories at any time.

- Our policy is not to accept or preserve personally identifiable information (PII), sensitive material, material subject to export control, or material that is primarily administrative rather than for research purposes. Our systems and staff attempt to intervene and prevent material with these features from being added for a combination of ethical, legal, and contractual reasons.
- If such content is discovered in our repositories, we will take steps to mitigate risks to the institution and its community, including removing the material causing this concern.
- In the eventuality that material is removed, we will make our best efforts to maintain a removal notice to which durable identifiers (eg, DOIs, Handle System URLs) will continue to resolve, and return the content we removed to the original creator.

How Long We Preserve Things

As noted above, while we can commit to preserving and providing access to some formats and content types indefinitely, there are some that require periodic retention reviews. These retention reviews may also result in decisions to de-accession older materials that we deem no longer relevant or useful. This is dependent on the resources that further preservation and access would require.

- All repositories strive to preserve content (i.e. type, not format) for the duration of that content's useful life.
- All the repository services will review their specific policies and the preservation worthiness of the content at least every 10 years. There may be cases in which more robust services are necessary for longer-term preservation, especially in cases where material will be retained for longer than 10 years.
- Generally speaking, we will move content among repositories, as needed, for the appropriate preservation and access of the material.
- Some content may be hosted on behalf of external partners according to the terms of a contract, and therefore removed from the repository at the request of the partner or when the contract's term ends.

Methods of Achieving Good Digital Preservation

Digital preservation is often discussed in terms of technologies and file formats, but the DRSSC considers preservation primarily a commitment of time, attention, and active management of data and metadata. These commitments mean we develop and maintain a variety of technological strategies and partnerships to achieve our goal of durability. As the future is unknown, successfully fulfilling our commitments requires a diversity of approaches to mitigate risk.

General Principles

The digital preservation community has identified a number of [basic strategies](#) for digital preservation:

- *Bit preservation*: Preserving the data exactly as it is, taking action to prevent data loss or corruption. This strategy is a prerequisite for all other digital preservation strategies; alone, however, it offers no promise of future usability or viability of the data due to file format obsolescence.
- *Migration*: Overcoming file format obsolescence by reformatting the content as necessary to usable successor formats. Choosing file formats that conform to open standards greatly increases the ability to do this without loss of content features; however, the migration process may result in changes to the presentation of the content.
- *Emulation*: Imitating the original (obsolete or unavailable) hardware or software on which the content was served in order to render the original digital object "as is", preserving the original presentation and user experience as much as possible.

We employ a number of methods to achieve bit and migration preservation strategies, as detailed below. U-M Library is a sustaining member of the [Software Preservation Network](#) and is following with interest research in the area of sustainable emulation strategies, but we are not yet employing an emulation strategy at this time.

The strength of our preservation commitment derives from its home within a Tier 1 research library which embodies preservation as core to its mission. As a result, we adhere to the [policies and practices of U-M Library](#), and content we host is preserved according to the same policies and practices, and with the same dedicated, committed effort, as all of U-M Library's digital materials.

[U-M Library Digital Repository Services Registered Formats and Support Levels](#) are based on the recommendations of the Library of Congress and reflect the accepted best practices of the digital preservation community. The service models of our individual repositories necessarily vary in the details of how these best practices are applied, while remaining true to the overall

strategy. In particular, the level of commitment to specific file formats may vary depending on compatibility with the purpose and goals of an individual repository, its technological capabilities, accessibility, and related publishing industry standards.

U-M Library's institutional stance is strongly predisposed to migration strategies. Our use of open, published standards puts us in the best position today to successfully migrate content to tomorrow's successor formats. On those occasions where no open standard is available, or the *de facto* standard within a given discipline is a proprietary file format, we may host the file with a commitment only to bit preservation (Level 3). In those cases, when the file becomes obsolete, we may continue to host it as-is as an artifact of possible historical interest.

Preservation decisions are based on a combination of best practices and the ability to provide the best possible access for a repository's Designated Community. Our access platforms serve content via repository software. Preservation files are managed in the U-M datacenter and are stored on and backed up using enterprise-quality equipment and services provided by U-M Information Technology Services. Strong linkages are maintained between preservation files and access derivatives, whether managed in the same or different systems.

Our Preservation Levels

Level 1: Things We Migrate

The Library will provide its highest level of preservation support, making its best effort to maintain the content, structure and functionality in the future. This service level is currently provided only for formats that are both publicly documented and widely used. The content may also be migrated (transformed to another stable format). Finally, the content will be preserved as originally deposited to ensure the original bitstream is always available. .CSV is an example of a Level 1-supported format, as its specifications are publicly available, it is well-supported and widely deployed.

These services include file format migration informed by ongoing file format monitoring, and the creation of preservation metadata documenting these transformative actions.

Level 2: Things We Try to Migrate

The Library will make limited efforts to maintain the usability of the file as well as preserving it as submitted (bit-level preservation). This level of support is generally applied to proprietary formats that are widely used and where there is substantial commercial interest in maintaining access to the format (e.g., Microsoft Excel). However, because of the uncertainty inherent in tool development for proprietary formats, the Library can only guarantee bit-level services and ongoing monitoring for level 2 formats at this time. We will migrate to new formats when both necessary and possible.

Level 3: Things We May or May Not Migrate

The Library provides basic preservation of the file (bitstream) and associated metadata as-is with no active effort made to monitor the format and associated risks or to normalize, transform, or migrate the file to another format. Files may be openable and/or readable by future applications, but there is no guarantee that the content, structure, or functionality will be preserved. 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. Shapefile (.shp) is an example of a format that would receive Level 3 support in the Library. Any format not yet reviewed and evaluated by the Library will also receive Level 3 service on deposit, but a higher level may be assigned after format review takes place.

All content will receive services commonly referred to as "bit-level" preservation. That includes the creation of archival backups, file format characterization, and protecting content from preservation threats such as bit-rot and unintended changes/deletions. All content will be stored on an enterprise-grade infrastructure with appropriate disaster recovery capabilities, security, and media replacement schedules.

See the complete list of the [U-M Library Digital Repository Services Registered Formats and Support Levels](#).

Preservation Partnerships

We fulfill our preservation commitments through a strategy of institutional and geographic diversity based on balancing local and distributed resources. By engaging in a number of such relationships our digital repository services guard against the risks of technological failure, changes in institutional priorities, or inability to meet commitments at any one institution, thereby mitigating single points of failure and ensuring the long-term access to content we host.

Our intent is to grow these partnerships for all our digital repository services. Partnerships will be initiated after a careful and thorough review of the benefits, costs, and responsibilities assigned to everyone involved. We will review partnerships on a regular basis to ensure that our expectations and needs for preserving our digital content are being met.