Managing Digital Assets

Preservation

Contents [hide]

1 University of Alabama Libraries Digital Services
- 1.1 Preservation Plan for Digital Materials
- 1.2 Division of Digital Content
- 1.3 Committed to sustain access
- 1.4 Migration Support
- 1.5 Emulation Support
- 1.6 Bit-Level Preservation
- 1.7 Annual Review
- 1.8 Long Term Retention
- 1.9 Local Backups
- 1.10 Descriptive Metadata

University of Alabama Libraries Digital Services

Digital content is inherently fragile. It is easily corrupted, damaged, changed, or deleted.

Hence, access to important digital content must be controlled. If what we are protecting is the historical record, change to the original content must be prevented.

Even running a virus checker across content can change it. Opening a file can change it. Moving a file from one media to another can change it.

How do we protect and preserve our unique, fragile, historical documents?

How do we make them accessible, both today, and in the future? The whole point of preservation is support of long-term access.

Incoming digital content adds another layer of issues to these two questions:

1. We may not know who or what has touched this content before we receive it, so we may be unable to guarantee its authenticity
2. It is likely not yet in archival format, and may not be of archival quality
3. It may not be in formats or on media with which we are familiar, or for which we have hardware or software
4. It may contain information that needs to be redacted or controlled, due to intellectual property rights, copyright, privacy issues, computer viruses, or other issues
5. It may contain information the donor did not intend for us to have
6. We may have little or no information about the content.

Read more about our processes for incoming digital content here: Managing_Incoming_Digital_Content

Preservation Plan for Digital Materials

The University of Alabama (UA) Libraries preserve selected digital content for long-term access support. Our highest level of attention and support is given to content selected for digitization from UA Libraries Special Collections. Other research materials are assigned preservation strategies at appropriate levels based on file formats and perceived needs of our designated audience, the faculty and students of the University of Alabama.

The University of Alabama (UA) Libraries DigiPres group will determine the need to normalize or migrate files pending loss of access due to obsolescence. Decisions will be made on a cost/benefit basis with consideration for the needs of our stated audience.

High Level Policies

Security and Business Continuity Policies

Recommendations for Authors and Creators

Division of Digital Content

1. **Level I support** is for content digitized in formats and with methods supporting the current archival standards, and for which we have digital rights management permissions and documented access permission. This is our most dedicated level of support. It includes collection of technical and administrative metadata, bit-level preservation, and commitment to migrate content as formats change over the years.An example would be a manuscript collection digitized by Digital Services.
2. **Level II support** is for content which may not have been digitized in currently supported archival formats, but for which The University of Alabama Libraries has committed long term access support, and for which we have digital rights permissions and documented access permissions. An example would be Electronic Theses and Dissertations.
3. **Level III support** is for content which needs to undergo regular change, and hence is not appropriate for inclusion in LOCKSS; however, it is to our benefit to offer bit-level preservation for this content until it needs to change. An example of this would be software necessary for either migration or emulation.
4. **Level IV support** is for content which may not have been digitized in currently supported archival formats, but for which The University of Alabama Libraries has committed short term access support, and for which we have digital rights permissions and documented access permissions. An example would be Undergraduate Research Papers.
5. **Level V support** is for content for which The University of Alabama Libraries has not committed access support, but which is currently managed by Digital Services, and for which we have digital rights permissions. An example would be files digitized at the patron request.

Support Level	Example	Committed to sustain access	Migration Support	Emulation Support	Long Term Retention	Bit-Level Preservation	Annual Review	Local Backups
Level I	Manuscript collection digitized by us	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Level II	Electronic Theses and Dissertations	Possibly	Possibly	Possibly	Yes	Yes	Yes	Yes
Level III	Open source software for rendering archival content	No	No	No	No	Yes	Yes	Yes
Level IV	Undergraduate Research Papers	No	No	No	No	No	Yes	Yes
	Material							

Level V	digitized at patron request	No	No	No	No	No	No	Yes

Committed to sustain access

Every feasible effort will be made to continue access to this content. This may involve migration to new formats, or development and maintenance of emulation methods. This level of institutional commitment can only be made for content created in current archival format standards. Content not created in current archival standards is much likely to be migratable to new formats. However, if the content continues to be of value and either such migration is feasible and retains the significant properties of the content, or if emulation support is feasible, then continued access will be supported.

Migration Support

1. Formats of archival files and versions of metadata will be stored on the top layer of the file system, in a flat text file exported regularly from the database where all entries to the storage system are entered and monitored regularly for format or metadata migration requirements.
2. Descriptive, administrative, and provenance metadata will be stored in current schemas and formats in the file system as specified.
3. Technical metadata will be extracted from archival files and formatted for storage into appropriate schemas (local profiles are currently under development, drawing from standards such as MIX for images, TextMD for text, AudioMD for audio).
4. Open-source software which renders the current archival format, if available, will be stored in the archive. This will enable migration to newer file formats after the current ones become obsolete.
5. A copy of an open-source operating system which supports the open-source software, if available and feasible, will be stored in the archive.
6. Software and documentation necessary for emulation (recreation of the current user experience of our delivery system) will be stored in the archive.
7. File system information which enables emulation of the operating system to support the file system will be stored with the content.

Emulation Support

1. In addition to the migration support above, open-source software needed for creating derivatives and providing web delivery may be stored in the archive.
2. Documentation of current procedures for recreating the current online user experience may be stored in the archive.

Bit-Level Preservation

1. MD5 checksum scripts will run before each tape backup to verify content is not corrupt, and will notify the repository administrator of any errors. Backup copies of current checksums are stored on a separate server, and scripts on a third separate server verify checking scripts run as scheduled and without error.
2. We are and will continue to be involved in LOCKSS or a similar preservation network, supporting at least 6 copies of the archival content across a geographically disbursed area. All archival content will be made available to this system.

Annual Review

1. Prior to obsolescence, all content will be evaluated for preservation measures, which may involve either migration (reformatting) or emulation. Dependent upon their decisions and the availability of resources and viable migration/emulation methods, efforts will be made to continue accessibility. All preservation measures taken will be recorded.
2. If continued accessibility is deemed infeasible or advised against, online access will end, and stored content and metadata will be deleted.

3. The definition of obsolescence used in these statements is that in which the approved computer systems and software on the University of Alabama Library computers can no longer render viable access to the content in the file without emulation services.

Long Term Retention

1. Digital content will be named according to our file naming scheme and organized according to our file storage scheme on our storage server.

Local Backups

1. The storage system is covered by a weekly full backup and daily differential backups. The weekly full backup is duplicated, and a copy is sent offsite, with at least a two-month rolling backup schedule.

Descriptive Metadata

1. Up to 2 versions of descriptive metadata will be stored, the original and the most recent. Captures will be made quarterly from the delivery software web directories. If the metadata found there is more recent than what is stored, it will be placed in the archive. Version 2 of each metadata file will be overwritten with each new capture.

Our current preservation Network is [The Alabama Digital Preservation Network]

- Organization of completed content for long-term storage
- Formats
- File Naming and Linking for LOCKSS

Jlderidder

NORTHWESTERN UNIVERSITY
Hydra. Digital Image Library
http://projecthydra.org/community-2-2/partners-and-more/northwestern-university-2/

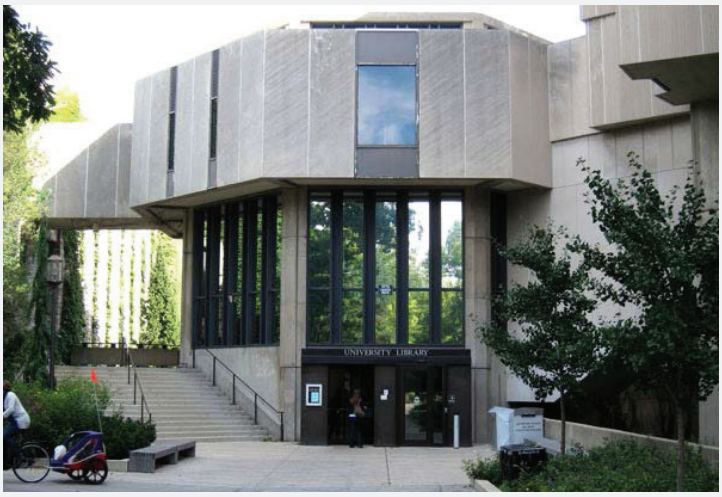
hydra

Get a head
on your repository.

Multi-Purpose Repository Solutions
Flexible User Interfaces
Durable Digital Asset Management

About | Applications & demos | Community | Design | Technical | News & events

Northwestern University



The library at Northwestern University

Digital Image Library

Northwestern University is working on a Hydra head that will provide extensive functionality for its Digital Image Library (DIL).

What is Digital Image Library?

DIL is an implementation of the Hydra technology that provides a repository solution for discovery of and access to images for staff, students, and scholars. It is comprised of three applications: a public-facing search and discovery interface, a restricted workflow management interface (currently in Drupal), and a restricted metadata editor interface (currently in XForms) with an Authority Tool feeding vocabularies from the library's integrated library system (Voyager).

Search

Search

News & Events

- Hydra winter meeting
- HydraHead 6.3.0 and ActiveFedora 6.4.0 released
- Virginia Tech becomes the 19th Hydra Partner
- Active Fedora 6.3.0, HydraHead 6.2.2, Sufia 2.0.1 released
- Avalon webinar

See All News & Events

Get Involved

- News & Events
- Hydra events diary
- Hydra wiki
- Communication channels
- Mail us

Developer links

- Hydra Wiki developers' pages
- Hydra on Github

Background

In 2007 the Library assumed responsibility for the Visual Media Collection (Art History Slide Library) and had to evaluate issues in three major areas of image collection management: digitization and cataloging, service assessment and promotion, and providing tools to make the collection useable. The Visual Media Collection was to become a wider Digital Image Library, thus collections and affiliated services were evaluated to determine how they would integrate with the digital repository and transition from serving one discipline to many. The result was the development of DIL : an institutional repository for images built using the Hydra technology framework and based on the Fedora digital repository system.

DIL is an implementation of the Hydra technology that provides a repository solution for discovery of and access to images for staff, students, and scholars. Some important features are:

- Build custom collection of images using drag-and-drop
- Re-order images within a collection using drag-and-drop
- Nest collections within other collections
- Create details/crops of images
- Zoom, rotate images
- Upload personal images
- Retrieve your own uploads and details from a collection
- Export a collection to a PowerPoint presentation
- Create a group of users and authorize access to your images

Our Hydra application also has a REST API that a robust metadata cataloging tool and a migration process utilize. This Hydra-based API ensures that the objects get indexed in Solr and updated in Fedora, all with simple REST calls.

Avalon

Northwestern University is collaborating with Indiana University on the Avalon Project. Avalon will use Hydra technologies to provide a powerful repository for use with video materials.

University website:

http://www.northwestern.edu/

Hydra website:

Digital Image Library (restricted to users with NU NetID)

Status: development

Screencast:

Digital Image Library (approx 3 minutes)

Digital Project Support Framework

Washington University Digital Implementation Group (DIG)

April 3, 2007

Revision J

1 INTRODUCTION

Recent years have seen the emergence of a number of scholarly digital projects on the Washington University campus. These have ranged from small student projects to larger faculty-driven undertakings such as American Lives. However, several obstacles to further development of such work remain, including long-term preservation, short-term support, a consistent knowledge base, common tool support, and integration of digital materials into larger digital library or repository systems. These problems often limit how these projects are valued as scholarly or pedagogical resources.

The primary purpose of this document is to establish a *lingua franca* for digital projects at Washington University, integrating the perspectives of faculty, library staff, and other interested parties in the University community. A significant step toward such a common understanding is the recognition of the challenges that different members of the community will face as they develop digital projects, and of our shared goals as we develop a University digital library and related infrastructure. This document does not attempt to establish *specific* procedures for accepting and developing digital projects, nor standards that such projects should follow. Instead, it represents an agreement as to what kinds of procedures and standards should be developed on a University-wide basis.

To that end, this document establishes different classes of digital projects as a preliminary step to providing them appropriate support. Explicit criteria as to what support any given project merits remain to be determined at a later date. Eventually decisions regarding the level of support allocated to a proposed project should be made on a consistent rather than an *ad hoc* basis. Furthermore, while the University Libraries have committed to playing a central role in providing such support, this document is not presented from the perspective of the Libraries, or any given school or division of the University, but rather from the University level, so digital projects created by the library would in no way be synonymous with "University projects" described in this document.

In conjunction with other institutional steps, this document also represents a commitment to provide a greater level of support to projects at all levels, and is therefore intended to increase development of digital projects, especially by faculty, and specifically to encourage development of digital projects as a scholarly activity. At the same time, it is intended to encourage this development in a disciplined way that will help to ensure the successful execution of digital projects, and to most effectively leverage the resources available for digital project development.

2 PROPOSAL PURPOSE

This paper describes a framework for handling digital projects at Washington University. The purpose of this framework is to address some of these issues and discuss ways in which the University can structure activities to support these projects.

The issues addressed are:

- Long-term maintenance of digital projects
- Role of a central digital library
- Role of a digital asset repository

3 PROJECT SCOPE - WHAT ARE DIGITAL PROJECTS?

For the purposes of this proposal, digital projects are defined as some combination of scholarly research, research tools, and collections of artifacts that are significantly computer-aided and usually web-based. For example, an interactive literary scholarly edition, a web site that presents an organized collection of digital photos and maps on twelfth century London, or a virtual exploration of the pyramids. What are not addressed by this proposal are interactive databases where the underlying content is expected to change rapidly or over long periods. For example, the student information system and the library catalogue are not covered. The focus, therefore, is on faculty or student-driven scholarly digital projects where the result is somewhat akin to a book, paper, or museum exhibit (in its formal intellectual content, not as media).

3.1 THE STRUCTURE OF A DIGITAL PROJECT

The conceptual structure of scholarly digital projects can be broken down into two general pieces.

1) Content— At the core of a digital project is the content made up of data and metadata. The data is the scholarly material. It may include images, film clips, paper or other text blocks, sound clips, maps, etc. Some of the material may be the work of the scholars involved in the digital project or it may be the work of others. The works may be digital in origin or digitized copies of non-digital work such as scanned images. Whereas the data is the primary scholarly information, the metadata describe information about the data. For example, the data might be a scanned photograph. The metadata might describe who took the photo, when it was taken, and when it was digitized. Metadata is the information needed to classify and catalogue the data. In theory, data with appropriate metadata could be incorporated into other digital archives.

2) Presentation—Presentation includes both tool development, which allows researchers to submit queries and derive specific information from a project's data set, and static presentation, such as the web page and interface of a project. So, for example, a literary archive may have a static web page through which users can call up different editions of an author's work; it may also allow users to pose queries, such as word counts within different documents. The web page is static and the querying tool is dynamic, but both are presentations of the content.

Scholars who wish to build digital projects must recognize the difference between content and presentation if they hope to develop projects that are responsive to research needs and are preservable for the long term. By properly creating data and metadata as separable from the tools and interface through which they are accessed, the content can be re-purposed (in part or in whole) and re-published in other formats, including future formats not yet developed.

In order to provide optimal support for digital projects, Washington University recognizes as a best practice the separation of content from presentation. Specific implementation of this best practice will vary from project to project, and will likely change in response to scholarly needs.

4 NON-PROJECT DIGITAL ASSETS

Not all digital projects properly belong to a digital collection or project. Sometimes members of the University community may create a digital object in isolation—a scanned photograph for classroom use, for example. In the analog past, personal collections of photographs would often be accompanied by clues that gave such objects context, such writing on the back identifying its subject or when or where it was taken. A significant drawback to digital resources is that they typically have little or none of this kind of identification. Typically, digital assets created for personal use in the classroom are only nominally identified, if at all.

Such assets become problematic when a faculty member approaches the university with curation or delivery requests. These classroom resources may constitute valuable resources that deserve preservation, but the lack of documentation for such a resource would present a significant obstacle to curation.

Washington University hopes to offer a curatorial service for these and other orphaned resources, or non-project digital assets, in the form of a digital asset repository, discussed later in this document. Such a repository will provide a valuable service to the university community, but will also require faculty and other creators of such assets to acknowledge minimal metadata and formatting standards in order to make their resources preservable.

5 WHAT ARE THE CHALLENGES?

Three interlocking challengest must be met for successful, long-term scholarly digital project development at the University.

1) **Duration** – Digital projects are created for various purposes, from limited short-term use in a single course to long-term, broader scholarly use. To complicate matters, the purpose of a project often changes over its lifetime. A project originally conceived as a tool in an individual's research may later be recognized as a valuable resource for an entire community. Finally, and most importantly, long-term preservation remains a stumbling block in the acceptance of projects as long-term investments. Unlike books, which stay fairly stable after publication, digital projects often die when the original creator retires, technology changes, or

when direct funding runs out. One of the goals of this framework is to propose a method to retain digital projects (or their contents) over decades, thus improving their value as scholarly work.

2) **Content (digital asset) management –** Content or digital asset management is important to the long-term success of the entire digital library endeavor. By properly segregating content from presentation—and even within these categories, separating data from metadata and static presentation from tool development—projects better ensure their longevity, and help clarify the roles of the scholars and curators involved. Once these pieces of a digital project are elucidated, it is easier for the library to ingest the data, and for scholars to study and share the resources across projects.

3) **Value as a scholarly activity –** Finally, digital projects and their contents present the same problem of recognized scholarly effort that any book or paper presents. How does one determine if a project is of scholarly value and should be preserved? There are established mechanisms in the print world for this evaluation. Peer-reviewed journals, book publication procedures, and library selection processes are all part of this process. Currently, similar mechanisms are not as codified in the digital world. Although this framework does not address the issue of scholarly value directly, it does maintain that the University must decide whether a project is worth long-term financial investment.

6 PROPOSAL

There are five elements to this proposed framework:

1. **Recognition of Presentation/Content Structure**
2. **Establishment of a Common Set of Project Definitions**
3. **Establishment of a University Digital Asset Repository**
4. **Establishment of a University Digital Library**
5. **Establishment of a Digital Project Web Portal**

6.1 Presentation/Content Structure

It is important to recognize a distinction between 1) developing and preserving digital content and 2) developing presentation and tools. This distinction will help clarify the responsibilities and investments required of various parties in the development of digital projects.

6.2 Common Set of Project Definitions

The following sections offer categories for describing a digital project's 1) support (divided into four classes), 2) approach to content, and 3) hosting.

6.2.1 Project Classes

A project's class defines how much support the school or University has commited to the project. If a school or the University commits significant support to a project, resources will need to be specifically allocated to the project. This proposal does not

determine how schools, the library, or the University will allocate these resources, since such decisions should be made by the school, library, or university itself.

- **Class 1 – Local Project**. No significant support from either the school or library. The project is completely controlled and developed by the local faculty or student groups. Funding may be from a department or external agency. Operation time length is up to the faculty or students.
- **Class 2 –School Supported Project**. Similar to Class 1 projects except there is significant support by the school. School supported projects will normally be required to meet standards set by the school.
- **Class 3 –University Supported Project.** Similar to Class 1 and 2 projects except there is significant support by the University (via the library and possibly the school). University supported projects will normally be required to meet standard set by the library and/or school.

6.2.2 Content Approach (Project Standards)

A project's content approach refers to whether a project implements standards that allow for data migration and preservation. Content approach can fall into three categories:

- **Type 1—**Local Use Only. In this content approach, data is created with no intention of having it preserved for the long-term or migrated to any third-party system, such as the University Digital Library or the Digital Asset Repository.
- **Type 2 –** Storage in Digital Asset Repository. Directors of a project using this approach would incorporate the minimal metadata and formatting requirements to enable the library to store their data in the Digital Asset Repository. The library would not be required to provide user-friendly interfaces, search functions, etc. for such data.
- **Type 3 –**Inclusion in the University Digital Library. The most labor-intensive content approach, this method incorporates enough metadata and otherwise responds to library requirements for ingestion into the Digital Library. The Digital Library provides at least a minimal infrastructure for retrieving data. Please note that meeting these standards does not guarantee ingestion into the Digital Library; it is simply a minimal requirement for acceptance.

Not all the content of a given project may fall into a single category. Some content may be generated at library archival standards for inclusion into the University Digital Library or Digital Asset Repository while other content may be generated just for use in the local project. Further, a Class 1, Class 2, or Class 3 project (as defined in §6.2.1) may be developed by a project team who plans to operate it for only a few years, but who hopes that the content will be curated for the long term. Thus, content for even a local project may be generated to meet library standards for future inclusion into the University digital

6.2.3 Presentation Approach

Adhering to metadata and formatting standards can help ensure the long-term preservation of a digital project's content, but a project's presentation is less durable. In fact, ongoing developments in data mining and analysis techniques virtually ensure that a given project's presentation will be updated continuously at a local level. Projects that invest in durable, preservable content provide the stable arena in which exploratory and innovative approaches to presentation become possible. Consequently, Washington University encourages projects to invest in durable, preservable content, and to view the upkeep of presentation as a built-in cost of digital projects.

6.2.4 Hosting

Hosting refers to what computer servers are used for the project. Servers, including backup systems, constitute a significant cost of a digital project. There are three general types of hosting:

- **Local hosting**–Hosted on local servers (project-specific, faculty, or student machines).
- **School hosting–**Hosted on school servers.
- **Library hosting–**Hosted on library servers.
- **External hosting**—Hosted on a server not sponsored by a Washington University entity.

6.3 Digital Asset Repository

Previous sections of this document have stressed the need for a Digital Asset Repository, which will hold standardized content that need not be part of any digital project or collection per se. Such a repository would preserve not only isolated digital objects created for classroom use, for example, but could also store stabilized data for projects that have less stable presentations. The Repository should, in other words, act as a clearinghouse for members of the university community who create digital content that meets the Repository's metadata and formatting standards.

The Digital Asset Repository will:

- Promote the use of standardized content
- Ease the problem of search and retrieval
- Promote the re-use of digital assets across multiple projects
- Ensure that even local (Class 1) projects have means to be preserved for the long term
- Encourage project developers to think in terms of content vs. presentation
- Help students quickly learn to design digital projects by providing them with pre-digitized content

The Digital Asset Repository will meet the needs of a wide range of the university community, from faculty and students creating single digital objects to larger research projects that would like to design their own presentations of their data while having the data housed elsewhere.

6.4 Digital Library

As part of their ongoing efforts to support education and scholarship, Washington University is building a core Digital Library that includes digitized versions of materials already held by the Libraries, some scholarly digital work created by faculty, and possibly other licensed resources. The central Digital Library will adopt an appropriate digital asset management system that stores collections, includes metadata that describe them, manages access to these collections, and facilitates delivery to users. It will present and provide access to content across many format types through a single web-based point-of-access site. A central digital production facility will ensure ease of digital creation workflow, conformity to accepted standards, and inclusion in the Digital Library. The University will adopt guidelines for digital projects and a procedure for the development of digital projects including intellectual property issues, creation of metadata, and production support.

6.5 Alternative Digital Services

The Digital Library will be an excellent resource for many members of the university community who are looking to digitize collections. It is important to note, though, that other options, such as building a project whose content is housed on an external server and whose presentation interface is housed on a school server, may better meet the scholarly needs of some researchers. These decisions are best made on a case by case basis after consulting with members of the digital community, such as Digital library Services or the Humanities Digital Workshop.

6.6 The Digital Project Web Portal

Finally, the University will create a digital project web portal that links to all sponsored digital projects on campus. The portal can also include important information such as policy documents and news announcements about digital work at the University.