Digital Preservation at UW–Madison

Report of the Long Term Digital Asset Management Working Group[1]

1. Executive summary

As more and more scholarly work is created in born-digital form, it is becoming critically important that institutions provide enterprise-scale, professionally managed services that can accept, maintain, preserve, and recover unique digital assets for an indefinite period. Funding agencies increasingly require researchers to complete data management plans, in which they identify and select publicly accessible repositories (hosted by institutions, disciplinary organizations, funders, or companies) not only to fulfill the public access requirements, but also to archive and preserve their research data. Consequently, the expectations of researchers for reproducibility and data reuse over time require that they have access to options for archival and preservation environments that include the ability to provide access to other researchers and to the public.

In June 2016, the Long Term Digital Asset Management (LDAM) Working Group, sponsored by the Vice Provost for Libraries and Vice Provost for Information Technology, was charged with

- determining the demand for digital preservation at UW–Madison,
- gauging the extent to which current campus services could meet that demand,
- and proposing processes through which campus may address any unmet needs.

The working group met to define digital preservation and its scope, then solicited and gathered use cases from representative campus entities using a survey and interviews. The survey was answered by twenty-five respondents who manage a wide range of content types, the majority of whom indicated that they were subject to federal funding public access requirements. Eight respondents agreed to a follow-up interview regarding their current practices and expectations for long-term data management.

While a majority of survey respondents defined "long term" as greater than ten years (up to "forever"), their data management tools varied widely: purchased off-the-shelf hardware, departmental infrastructure, DoIT storage and archiving services, and institutional or disciplinary repositories.

---

[1] Endorsed by the UW–Madison Vice Provost for Libraries/University Librarian and Chief Information Officer, July 2019

When asked what should be included in a long-term digital management service, respondents selected the following services (in decreasing rank of selection):

- Curation: maintaining, preserving and adding value to digital data throughout its lifecycle[2]
- Maintenance of the software used to access the content: discontinued software, protocols, standards, format migrations, etc.
- Periodic integrity checking.

Respondents recognized that long-term management involves services that are layered above basic storage (which was consistently ranked lower).

Interviewees had differing opinions on the desired economic model for long-term asset management. While all supported some level of shared funding, groups relying on grant funding tended to favor a one-time, up-front charge, while those with a departmental budget preferred an annual payment model.

UW–Madison, through DoIT and its departments, offers a range of services relating to long-term digital asset management. Most of those services comprise various tiers of storage and backup, though groups like Research Data Services also provide consultation on data curation and management plans. The LDAM group did not identify any campus service currently providing true digital preservation. However, the Libraries are in the process of developing a preservation platform for their digital collections, leveraging campus infrastructure and local curation and management. Additionally, some researchers departments participate in external or consortial initiatives that provide preservation services. This landscape to some extent mirrors that of our peer institutions, though a small number have launched major initiatives to preserve digital assets at scale.

From our research and discussions with campus units, it seems clear that the value of UW–Madison's digital assets (measured both in economic and cultural terms) cannot be correlated with creators' ability to pay for its management over time. It is important, therefore, that the campus develops a service model that includes a common-good solution for long-term archiving or full preservation for high-value content. Given the cost of full preservation, decision matrices need to take into account a range of cultural, financial, technical, and regulatory factors, balancing 1) the perceived value

[2] Digital Curation Center, "What is Digital Curation?" http://www.dcc.ac.uk/digital-curation/what-digital-curation

of the digital asset, 2) the long-term risk to its integrity and interpretability, and 3) the cost of archival or preservation management.

The LDAM Working Group recommends that a process be initiated to:

- Articulate clearly to the campus community the case for digital preservation, and the organizational, technological, and financial resources needed to pursue it.
- Create tools and resources to help creators and administrators evaluate value and risks for their digital content.
- Develop archival and preservation service models.
- Develop a deeper inventory of existing campus services, and explore the most effective interaction and collaboration among those services
- Investigate potential business and funding models for archival and preservation services, including central, shared, client, and external funding.

This work should be coordinated with appropriate campus IT governance groups, and may involve the creation of one or more task forces or working groups.

2. Introduction and background to the report

This report is produced by the Long Term Digital Asset Management Working Group, formed in 2016 for the purpose of examining the state of digital preservation on campus, identifying gaps, and recommending next steps.[3] The workgroup is composed of librarians and DoIT IT staff experts in the areas of storage, preservation, data management, and data curation.

The working group first met to discuss in broad outlines what the scope of a digital preservation service should be, who were to be likely stakeholders, and what questions needed to be considered when contemplating such a service. It then solicited input concerning preservation needs from a diverse set of campus entities, chosen to represent the spectrum of kinds of digital assets being created and managed on campus. The results of these interviews are summarized below.

Armed with a rough set of needs articulated in the interviews, the working group then inventoried the services campus currently offers to meet those digital preservation

[3] "University of Wisconsin–Madison Long-Term Digital Asset Management (LDAM) Working Group Charter", https://uwmadison.box.com/s/4jmd8r45ctx1ijwst5amnt16h2oa00vi

needs, and investigated what peer institutions were doing in this area. The workgroup subsequently began to create a list of gaps and requirements. That work is still ongoing.

This report presents a brief case for digital preservation, based on the information collected to date. Digital preservation is defined and context is provided for the environmental trends that are driving memory institutions to commit to sustainable digital preservation programs are briefly covered. Current campus needs and services are then described, followed by the high-level environmental survey of peer institutions. Finally, next steps are recommended.

The audience for this report is campus administrators responsible at a high level for the custodianship of digital assets being produced under their purview. It assumes a familiarity with information technology and an abstract understanding of data management concepts.

3. What is Digital Preservation?

As the working group met with campus stakeholders, it became clear that it would be important to define terms carefully. Much of the terminology in this area is open to varying interpretations: "archiving", "backup", "storage", and "preservation" were often used interchangeably to refer to content management strategies.

The Association for Library Collections & Technical Services (ALCTS) Working Group on Defining Digital Preservation has developed short, medium, and long-format definitions for Preservation. The medium definition reads:

> "Digital preservation combines policies, strategies and actions to ensure access to reformatted and born digital content regardless of the challenges of media failure and technological change. The goal of digital preservation is the accurate rendering of authenticated content over time."[4]

In this light, *preservation* includes much more than the related concepts of "backup" or "archiving"; it implies more active management and planning, over longer time scales. It goes beyond purely technical strategies (which include backup and archiving) to include administrative and social factors. For example, the ALCTS group's long definition mentions the need for policies to "document an organization's commitment to preserve

[4] ALCTS Preservation and Reformatting Section, Working Group on Defining Digital Preservation. "Definitions of Digital Preservation" June 24, 2007 <URL: http://www.ala.org/alcts/resources/preserv/defdigpres0408>

digital content for future use". Preservation policies, practices, and services must be developed with that commitment in mind.

A number of units on campus provide services in support of the creation and long-term maintenance of digital archives, including the Libraries and DoIT. While these services address important technical needs like data curation, backup, secure storage, and long-term archiving, the LDAM Working Group did not identify any campus service that currently meets the definition of true digital preservation, which is realized as a holistic framework that includes commitment, long-term strategy, and coherent policy.

4. Environmental Trends

Increase in the amount of data being generated by researchers.

Many researchers are now regularly working with data on the scale of terabytes. The pressures for storage, both for active storage and long-term preservation solutions, are only increasing with the growing interest around big data and data science methods.

Changes in data sharing and preservation expectations.

Research culture and expectations are changing. As noted, there is more data and a shift in expectations for sharing and preservation from funding agencies, but there is also a growing awareness around issues of reproducibility of research and reuse of data. After studies completed by the Reproducibility Project from Center for Open Science showed that literature in the studied fields was largely unreproducible, there has been a movement towards increasing the transparency, openness, and sharing of research processes and outputs. There is also a growing awareness around the need for data to be available for reuse and reproducibility for other reasons, as demonstrated in the case of Zika virus. UW–Madison researcher David O'Connor shares his Zika research openly and immediately due to the global health implications as well as due to ethical implications of non-human primate research. Likewise for expensive, hard to reproduce, or longitudinal research, sharing data for reuse provides benefits to other researchers who may not have access to the same resources. Data sharing also encourages outside inquiry and peer review. These issues and changes in expectations of researchers around reproducibility and reuse over time require that researchers have access to options for preservation and storage environments that include the ability to provide access to researchers both on and off campus as well as to the public.

Funding agency requirements

Following the 2013 Office of Science and Technology Policy memo, *Expanding Public Access to the Results of Federally Funded Research*, federal agencies with over $100 million in R&D were required to create plans for public access to the publications and data resulting from grant funding. While the publication sharing requirement has largely been clarified and agencies have identified specific repositories in which to share that content, the direction for research data has been less direct. Agencies ask researchers to complete data management plans, in which they identify and select publicly accessible repositories (hosted by institutions, disciplinary organizations, funders, or companies) not only to fulfill the public access requirements, but also to archive and preserve their research data. Depending on the formats, size, and types of data researchers are producing, selecting an appropriate repository solution can be confusing, difficult, or costly. Recently, funding agencies have started to provide further guidance on where researchers should be putting their data long-term. The NSF now includes the phrase 'institutionally supported repositories' in their data guidance, and it is likely other agencies will follow.

5. Current campus Environment

Summary of campus survey

The working group administered a survey designed to assess the demand for digital preservation services and to understand the landscape of needs on the campus. The survey received twenty-five total responses. Eleven of the respondents said they are the creators of or are directly responsible for the digital assets, while nine respondents said they are in positions of support for others who create digital content. The rest selected 'other' or did not answer.

Of the respondents, all but two felt that their content had long-term value. Processed or analyzed data was the most common type of digital asset of those surveyed; however, the respondents had a fairly even distribution of assets across all of the provided content types (options provided were: Cultural Artifacts, Raw Instrument Data, Processed or Analyzed Data, Administrative Records, and Publications). Of the respondents that selected 'Other', the types of assets they listed included numerical models, aggregated research data, and educational content. Estimates of current

content that *should* be preserved varied: of twenty responses, five answered less than one terabyte, nine answered one to ten terabytes, and six answered over ten terabytes (going up to 100 terabytes). Estimates of new content created annually were lower, with ten respondents answering less than one terabyte, eight answering one to ten terabytes, and two answering over ten terabytes (up to 500 terabytes).

When asked what regulations or mandates may affect the preservation decisions for content, the majority of respondents answered that they were subject to federal funding public access requirements. Other regulations included state record retention guidelines, HIPAA/FERPA/IRB guidelines, and collections policies. Seventeen of the respondents answered that their content needed to be readily available to granting agencies, collaborators, and the public. Grant cycles seemed to have varied impact, with fairly even distribution across answers ranging from no impact to high impact.

Current long-term management methods employed by respondents varied. Answers included using local or shared servers, purchasing hard drives and other hardware, using DoIT storage and backup solutions, tape storage, depositing content in repositories (institutional, disciplinary, content agnostic, and federal archives), sharing via personal websites, and exploring digital asset management (DAM) software. Some respondents indicated that they were just beginning discussions or that individual investigators were responsible for their own solutions.

When asked to define 'long term', the majority of respondents answered either eleven to fifty years or forever. Three respondents answered six to ten years, three selected 'Other' which they defined as the following (paraphrased) - permanently, hundred of years for publications and tens for data, and a minimum of three for some data but as long as possible during continued research.

Responses to the question about the impact of the current campus assumption that individuals are responsible for digital asset management again varied. Some respondents felt they had good solutions, but had other issues such as organization, sharing, and continual increase in need. Some noted that they liked having control over the solution they choose to implement, but that having a central solution may provide stability and standardization. Some noted that they are very much driven by cost issues and some indicated implementing ad-hoc solutions for their assets. When asked whether the University should provide a campus solution, fourteen respondents indicated positively or said 'yes'. Those who said 'yes' suggested the cost be on the order of a few cents to a dollar per gigabyte, be free up to a certain number of gigabytes, or be competitive with other available solutions. One respondent indicated that they thought it should be provided for researchers out of the university overhead

from grant funding. For those respondents who indicated 'no', some said that more resources in the form of policies, staff, or consultants who could help guide this work would be more beneficial.

Respondents selected the following services when asked what should be included in a long-term digital management service (in order from most selected to least selected):

- *Curation. Services to assist in the selection, maintenance, and enhancement (e.g.metadata) of digital assets for the purpose of long-term preservation. (14 respondents)*
- *Maintenance of the software used to access the content. Even if the content is preserved,it may not be usable without the proper software. (11 respondents)*
- *Periodic integrity checking. Processes and systems to ensure that the preserved content remains the same over time. (11 respondents)*
- *Generic geographically distributed storage. Two or more copies persisted in geographically distributed storage systems. (10 respondents)*
- *Generic replicated local storage. Two or more copies persisted in one or more UW–Madison managed storage systems. (7 respondents)*
- *Generic local storage. A single copy persisted in a UW–Madison managed storage system. (6 respondents)*

Summary of campus interviews

The Working Group interviewed community members to learn more about individual units' concerns and approaches to digital preservation at the University of Wisconsin–Madison. Interviews were conducted with representatives from eight different units from across the campus, representing a range of administrative and research perspectives. The following summary reports the key observations, common themes, and a range of perceptions and practices.

Storage and backup solutions used by interviewees ranged widely, from DoIT-hosted storage, shared local drives, collaborative cloud platforms like Box, and external platforms such as the Open Science Framework. Interviewees shared that these solutions often felt incomplete and that they did not have a formalized workflow. The interviewees also commented that the amount of time required to oversee the technology needs of their unit took away from other job responsibilities. Many were all too aware that their unit's storage solution did not equate to preservation and they sought to ensure that their materials were accessible in perpetuity.

It was important to the working group to understand how available our interviewees thought the preserved materials should be. Several researchers preferred that their materials be fully open to the public and broadly accessible, citing both their commitment to research transparency and the need to comply with federal funding mandates requiring public access to publications and underlying data. The researchers we interviewed had funding from several major federal agencies, including the NIH, NSF, NASA, and NOAA. Most of the researchers noted that some of their material could be shared openly but other material could not or should not be openly available. The reasons for this included privacy concerns or sensitivity of the materials, institutional ownership (as in the case of marketing materials), or because the materials were already widely available elsewhere. One interviewee mentioned that he considered a preservation service a "safety box" where materials could be deposited without fanfare or provisioning access.

When asked how large their files were, several units cited overall data collections between 1-5TB, another cited up to 90TB in its overall holdings, and one unit generates 50-100TB per year. Depending on the rate of file accumulation, it was easier for some units to consider yearly accumulation of files, whereas for others it was easier to cite what they have to date. Due to the range of answers and different measurements, more work would need to be done to gain a better estimate of the amount of data that needs to be preserved by units across campus. Units tasked with collecting cultural heritage materials, tended to have larger holdings, as did large, well-funded research centers that produce sizable quantities of data.

Interviewees had differing opinions on the desired economic model for long-term asset management. While all supported some level of shared funding, groups relying on grant funding tended to favor a one-time, up-front charge, while those with a departmental budget preferred an annual payment model.

Inventory of services/technologies currently in use on campus

The following services on campus provide some aspects of storage and preservation management in their offerings.

Campus Computing Infrastructure (CCI) - Physical and Virtual Server Hosting and Data Storage

Advanced Computing Initiative - Campus Research Computing Resources – Center for High Throughput Computing (CHTC) and CloudLab

MINDS@UW: University of Wisconsin's institutional repository, designed to gather and preserve scholarly outputs of the University's research and teaching mission. Users self-submit their content to the repository in consultation with the repository manager. MINDS@UW is not appropriate for institutional, administrative, or business records.

UW Digital Collections: The University of Wisconsin Digital Collections (UWDC) works collaboratively with UW faculty, staff, and librarians to curate and provide access to digital resources that support the teaching and research needs of the university community, uniquely document the university and State of Wisconsin, and provide access to rare or fragile items of broad research value. The UW–Madison Libraries are committed to managing these collections, currently comprising over two and a half million digital objects, far into the future.

Hitachi Content Platform - DoIT and Library Data Archive Service

Bucky Backup - Bucky Backup Lite (Disk based) and Bucky Backup Archive (Tape based)

UW Cloud Services - Amazon Web Services, Microsoft Azure, Google Cloud Platform, IBM SoftLayer

UW Cloud File Storage – Box, Google Apps, Office 365

Recovering Analog and Digital Data (RADD): a collection of hardware and software in the Information School Library designed for digitizing at-risk analog materials as well as recovering data from at-risk digital media.

LabArchives - Electronic lab notebook software, a research and data management and storage solution for researchers.

Research Data Services - An interdisciplinary organization committed to advancing research data management practice on the UW–Madison campus. RDS offers three main services: data management plan assistance, consultations, and education and training. RDS focuses on providing researchers with the tools and resources that support their efforts to store, analyze, and share data.

Departmental IT Services – Social Science Computing Cooperative (SSCC), Institute for Clinical and Translational Research (ICTR), DiscoverIT Storage, Computer Systems Lab, Department of Physics, Biochemistry, Bock Labs

Consortial initiatives

HathiTrust: UW–Madison is a founding member of HathiTrust, a partnership of major research institutions and libraries working to ensure that the cultural record is preserved and accessible. The mission of HathiTrust is to contribute to research, scholarship, and the common good by collaboratively collecting, organizing, preserving, communicating, and sharing the record of human knowledge. To date, over 560,000 of HathiTrust's 16+ million volumes are drawn from the collections of the UW–Madison Libraries.

Digital Preservation Network: As an institutional member of DPN, UW–Madison has recently signed a deposit agreement that will allow content to be geographically distributed across several preservation repositories for a minimum of 20 years. The Libraries are currently working with DPN to create a workflow for the deposit of UW Digital Collections content.[5]

6. What peers are doing

Overview

Many US academic research institutions of similar size and profile as UW–Madison have digital preservation programs in various stages of development, managed by the campus libraries, in collaboration with campus IT. Of the Big Ten Academic Alliance, only one institution offers preservation services to campus: Indiana University. A few other library systems offer consultation services to campus, generally as part of their research data outreach services.

With the exception of Indiana University, most digital preservation efforts that are documented appear to be at the National Digital Stewardship Alliance's (NDSA) Level 1 and Level 2 standards[6]: data is backed up, sometimes stored offsite or in multiple geographic locations, and data corruption may be monitored with file fixity checks.

---

[5] The Digital Preservation Network ceased operations in December 2018.

[6] Phillips, Megan, *et al. The NDSA Levels of Digital Preservation: An Explanation and Uses* http://www.digitalpreservation.gov:8081/ndsa/working_groups/documents/NDSA_Levels_Archiving_2013.pdf

The vast majority of institutions have at most one staff member, employed by the library, dedicated to digital preservation and data curation. Most libraries only do in-house digital preservation, usually under the umbrella of digital collections, preservation, archives, or special collections departments. Of the few libraries that do manage digital preservation as a separate activity, it is generally staffed by two or three librarians with expertise in the fields of digital scholarship, data curation, and research data management. To the extent that any institution offers digital preservation services outside the walls of the library at all, it is always in conjunction with other services: institutional repository, data curation, research data management, research discovery, etc.

In the few cases where digital preservation is offered as a broader campus service (or where such a service is contemplated), management of the service is generally made up of working groups that comprise primarily librarians and a few campus IT professionals, with some representation from interested stakeholders in other departments.

Big Ten Academic Alliance

Of the Big Ten Academic Alliance member institutions, the following have digital preservation programs, all managed under the auspices of their libraries:

- University of Illinois at Urbana – Champaign: preservation of library-curated materials
- Indiana University: campus-wide effort, with more than 80 departments contributing materials, with a concentration on cultural and memory institution artifacts and endangered media.
- University of Iowa: preservation of library-curated materials; significant investment in third-party digital preservation services (Archive-It, Portico, LOCKSS, CLOCKSS, DPN)
- University of Maryland: no digital preservation program per se in place yet, but they do group a number of related activities under the umbrella of digital stewardship. Focus is on institutional repository, library-curated materials, and format conversion for in-house library use.

- University of Michigan: preservation of library-curated materials and consultation services to departments
- Michigan State University: preservation of library-curated materials and consultation services to departments
- University of Minnesota: preservation of library-curated materials, with the potential to extend the services across campus
- University of Nebraska – Lincoln: mixture of data repository, institutional repository, and digital collection (on the ExLibris Rosetta platform) services, under a broad umbrella of preservation, but no formal preservation strategy that crosses these areas
- Northwestern University: preservation of library-curated materials, under the purview of the preservation department. Limited in scope.
- Pennsylvania State University: preservation of library-curated materials, under the purview of the preservation department. Limited in scope.
- Purdue University: preservation of library-curated materials, under the purview of their archives and special collections departments. Focus is on digital forensics, the program partners with the Purdue University Research Repository, a service for research data publication and discovery
- Rutgers University: in collaboration with their Digital Curation Research Center, a lab and workspace for pursuing digital curation activities, Rutgers also offers RUcore, a suite of institutional and data repository services, focused primarily on search-and-discovery.

The most ambitious project is at Indiana University, with over thirty staff members involved in preservation activities across ten working groups. The institution closest to UW–Madison in terms of size, scope, and ambition is the University of Minnesota. The Digital Scholarship and Curation program at Michigan State University also has similar goals, albeit smaller in scale and covering more areas, with five staff devoted to digital preservation, digital scholarship, and research data services.

Additionally, all BTAA institutions are members of HathiTrust, a consortium formed to manage and preserve the content of millions of volumes digitized from member libraries' collections.

Other US Institutions of note

As noted above, most digital preservation services currently are scoped to library content, with little or no engagement with data and materials outside library collections (either actual or potential).

The highest-profile exception to this model is the Research Data Curation Program at UC San Diego, a campus service that supports research data management, data search-and-discovery, and digital preservation. The Research Data Curation Program is also responsible for managing Chronopolis, a Digital Preservation Network node. The program is managed with a staff of eight librarians and technologists, in collaboration with departments across campus and other institutions.

The University of Washington ResearchWorks Service provides faculty, researchers and students with tools to archive and/or publish the products of research including data sets, monographs, images, journal articles and technical reports. It combines traditional search-and-discovery institutional repository services with research data management and limited long-term digital artifact preservation. They also manage online journal hosting and geospatial data archives under its umbrella.

The Carolina Digital Repository at the University of North Carolina – Chapel Hill is an institutional repository for the university community that also encompasses campus digital collections and research datasets as well as more traditional repository materials. It is notable for its widespread acceptance across campus: 162 schools, departments, research centers, and affiliated institutions make use of its services.

7. Conclusions and Next Steps

Conclusions

It is not surprising that the data curation, management, archiving, and preservation needs of UW–Madison departments and researchers vary considerably, both in the scope of activities needed and in the researchers' abilities to fund and/or manage those activities. As detailed above, a number of existing campus services are available to meet a wide range of use cases, particularly for storage, backup, and medium- to long-term archiving. To date, interaction between these and other interested parties has been mostly "one off", or project-based.

While many campus units are taking advantage of existing services, others find it more cost-effective to do it for themselves, at varying levels of sophistication. Still others lack

both the funding to use campus infrastructure and the expertise to create their own, and therefore rely on whatever methods they can patch together to meet their minimal requirements. The Working Group did not identify any campus unit currently supporting full digital preservation, although the Libraries have begun the process of developing a preservation repository.

Because the value of digital assets (measured both in economic and cultural terms) cannot be correlated with creators' ability to pay for its management over time, it is important for UW–Madison to develop a service model that includes a baseline common-good (*i.e.*, centrally funded) solution for long-term archiving or full preservation for high-value content, at least for those data creators without the ability to fund long-term management.

The development of tiered service offerings, from basic storage to full digital preservation, necessitates the creation of an evaluation process to help data curators and administrators determine the level of management required for a given digital resource over time. Given the cost of full preservation, decision matrices need to take into account a range of cultural, financial, technical, and regulatory factors[7], balancing 1) the perceived value of the digital asset, 2) the long-term risk to its integrity and interpretability, and 3) the cost of archival or preservation management. At the very least, a clear articulation of these factors will help to create a shared understanding across campus of both the risks and opportunities of digital asset management, and give the UW–Madison community a shared vocabulary with which to discuss these issues.

The unique, irreplaceable products of university activity increasingly only exist in digital form. However, there does not currently exist a framework and processes for protecting these digital assets for future generations that can match the systems in place to protect the university's more tangible assets. We believe there is value in bringing together the current providers of digital asset management services and those with content in need of long-term care, to develop a vision and strategy that will meet the current and future digital preservation needs of the University, on a campus scale.

Next Steps

The LDAM Working Group recommends that a process be initiated to:

[7] See, for instance, Currall, James, and Peter McKinney. "Investing in Value: A Perspective on Digital Preservation" D-Lib Magazine 12:4 (April 2006). doi:10.1045/april2006-mckinney

- Articulate clearly to the campus community what digital preservation is, when and why it is necessary, and the organizational, technological, and financial resources needed to pursue it
- Create tools and resources to help creators and administrators evaluate value and risks for their digital content
- Develop archival and preservation service models, defining increasing levels of preservation activity and the services (including existing ones) offered at each level
- Develop a deeper inventory of existing campus services, and explore the most effective interaction and collaboration among those services
- Investigate potential business and funding models for archival and preservation services, including central, shared, client, and external funding

This work should be coordinated with appropriate campus IT governance groups, and may involve the creation of one or more task forces or working groups.

LDAM Working Group

Sponsors:

Ed Van Gemert (Vice Provost for Libraries and University Librarian)

Bruce Maas / Mike Lehman (Vice Provost for Information Technology and CIO)

Lee Konrad (Associate University Librarian for Digital Strategy)

Steve Krogull (Director of Systems Engineering and Operations, DoIT)

Working Group:

Cameron Cook, Libraries

Steve Devoti, DoIT Enterprise Architect (co-chair)

Peter Gorman, Libraries (co-chair)

Kevin Kettner, DoIT SEO

Mike Layde, DoIT SEO/CCI

Scott Prater, Libraries

Past members:

Luke Bluma, DoIT, SEO/CCI

Josh Harder, DoIT Academic Technology

Brianna Marshall, Libraries

Appendix A: Core Characteristics of a Digital Preservation Service

Policy and Administration

The following are the administrative and policy-related activities and concerns in a digital preservation service.

Operational Administration

Tasks and workflows that support the day-to-day operation of the archive.

Planning

Processes and systems that look forward to ensure the goals identified for the archive are realized.

Succession

Policies that Identify and document how ownership and management of content may be transitioned over time.

Curation

Activities to normalize and enhance content to make it useful for future generations, as well as to facilitate its ongoing preservation.

Community-facing Services and Activities

The following are the high-level services or collaborative activities needed to support users, facilitate content deposit, and raise awareness of the digital preservation service.

Consulting

Resources to help consumers discover and select archive services and prepare their content for the archive.

Marketing and Communications

Programs and outreach designed to raise awareness of at-risk content and the services available to users.

Education and Training

Classes and other resources to enable consumers to effectively use archive services.

Partnering

Alignment of the various campus and external resources and groups to meet the needs of campus digital preservation service users.

Technical Environment

The following are the high-level technical infrastructure systems that form the core of a digital preservation service.

Storage

Physical systems where the archived content is stored.

Ingest and Access Systems

Workflows that allow authorized users to submit content to be archived, as well as retrieve archived content.

Data aggregation and organization

Processes and systems that group and/or associate data about the content with the content itself to ensure it can be managed and discovered over time.

Technology Monitoring Systems

Technology obsolescence, both physical and software pose a major risk to the long-term viability of the archive. Monitoring allows the archive administrator time to plan for format, software and physical storage migrations.

Content Monitoring Systems

Ongoing content monitoring ensures that 1) the content maintains its integrity; 2) that it can be identified as at-risk of obsolescence, and 3) that evidence can be provided showing compliance with the repository's digital preservation policies and obligations.

Compliance

Compliance is made up of the policies, standards, regulatory and legal mandates, and best practices that ensure that the digital preservation service meets its obligations in a way that is both correct and verifiable. Compliance includes, but is not limited to, the following:

Information Security

Policies, processes and systems that maintain the confidentiality, integrity and availability of the archive.

Policy and Governance

Structures for ensuring that the archive complies with University and other applicable policies or regulations. Policies developed within a governance structure support the creation and monitoring of archive goals and objectives. Governance provides a a decision-making body to solicit stakeholder input and hold the digital preservation service accountable, as well as providing a venue and process to develop long-range strategy and maintain continuity.

Appendix B: Definitions

Digital Preservation: "The active management of digital content over time to ensure ongoing access" (American Library Association/ALCTS)

Federation: Agreements and technology that connect consumers and providers of content when they exist in different organizations. Supports distributed access and replication.

Format Migration: The conversion of content to new formats as necessary in response to new software versions, new standards, or the deprecation of older formats, to ensure the continued usability of the preserved content.

Ingest: The process that actually takes content and persists it in the archive. Ingest applies appropriate metadata, including content descriptions, identifiers, as well as formats and other information useful in discovery and using the content. The ingest process also puts in place the metadata necessary to ensure the authenticity of the content being preserved.

Integrity/Fixity: Integrity/fixity are the means used to ensure that content does not degrade or become corrupted over time.

Metadata: Data about the content being preserved, which is used to 1) provide context for future analysis and reuse of the content, and 2) enable preservation activities on the content over time.

Replication: The act of making multiple copies of a resource in disparate locations, to protect content from both technical and physical risks.

Search and Discovery: The means to find content of use to a consumer in the repository.

Software Migration: The act of changing software systems used to view and manipulate content to keep up with current versions and standards. This usually goes hand-in-hand with content **format migration**.

Appendix C: LDAM Survey Questions

Detailed results are available upon request.

1. Do you have content that you believe has long-term value?

2. What is the nature of the content that might merit archiving or preserving for the future?

3. What steps, if any, have you taken to provide for the long-term management of your digital content?

4. What regulations, policies, guidelines, mandates, etc. are you aware of that could affect how and when you archive or preserve digital content?

5. Does your content need to be readily available to granting agencies, collaborators or the general public?

6. What does "long-term" mean to you?

7. Assuming not all digital content needs to be preserved, estimate the amount of digital content you currently have and the amount you will generate that should be archived or preserved.

8. What is the impact of grant cycles on your actions related to preservation of digital content?

9. Right now campus groups are responsible for their own long-term digital asset management. How does this impact the quality of the implementation of your digital asset plan?

10. Do you believe the University should offer a long-term digital asset management service to campus? If so, do you have ideas on how that would be funded given the current budget climate? What would be a reasonable cost for such a service, e.g. per gigabyte?

11. What should be included in a long-term digital management service?

12. Would you be willing to be interviewed about your digital archiving and preservation needs?

13. What is your primary interest in preservation?

Appendix D: LDAM In-person Interview questions

Detailed interview responses are available upon request.

1. Please tell us a bit more about yourself, your role, your organization. (5 minutes)

2. Please tell us more about the digital materials you are considering preserving. (15 minutes)
   - What value does it have that makes it a candidate for preservation?
   - What aspects of the materials, information about the material, should be preserved?

3. In your mind, what constitutes preservation? (20 minutes)
   - How long?
   - How many copies, where?
   - How accessible should the preserved data be? To whom?
   - How important is it that the materials be migrated over time to keep up with changes in formats and tools?
   - What are some of the potential uses you see for this preserved data, where some person, group, organization or robot in the future may want to recover these materials for some purpose?

4. Tell us more about any regulations, policies, guidelines, mandates, etc. that influence your preservation decisions and activities. (10 minutes)

5. Preservation has costs. Tell us how you think about the resources (people, storage, etc.) required to prepare materials for preservation, store them, and manage them over time. (10 minutes)
   - How much does cost (in terms of money, time, and effort) influence your preservation decisions?
   - Would a campus-wide preservation service be of use to you? If so, what might that look like?
   - Assuming that such a service is unlikely to be 100% centrally-funded, what kind of cost models for such a service would work for you?