Creating a Foundation for Digital Asset Management and Preservation

Digital Asset Management and Preservation Working Group (DAMP!)
September 2015
Rachel Appel, Kate Carter, Susan Dreher, Anna Goslen, Amy Mayer, Pat O'Donnell, Krista Oldham, Vince Patone

Executive Summary
Background
Current Landscape: Limitations and Opportunities
Decision Criteria
Systems Considered
Recommendations
Project Overview
Strategy

Appendices
Project Overview Timeline
Literature Review
Environmental Scan of Peer Institutions
An Abridged History of Digital Assets at the Tri-College Libraries
Content Map
Preservation Assessment of Current Systems
Generic Workflow
Glossary
Executive Summary

Digital asset management is the systematic process of acquisition, description, discovery, and retrieval of collections of digital objects in a streamlined, efficient, and sustainable manner. Digital asset management frequently includes the act of digital preservation, which is the maintenance and management of digital objects, including both those that are born-digital and those that were converted from analog formats, so that they can be accessed and used in the future despite anticipated technological obsolescence and unanticipated occurrences of data loss. A digital asset management system, or DAMS, allows an institution to manage, organize, and distribute its digital assets from a central repository.

As the Tri-College Library Consortium (TriCo) has been digitizing collections and acquiring born-digital materials over the last 15 years, our ability to properly manage digital assets according to best practices has been hindered by a spike in volume and a lack of a flexible DAMS. We currently have six primary DAMS in order to address the variety of objects, collections, and projects we facilitate. This multitude of systems has consequently fragmented our ability to properly manage our collections. The infrastructure, or lack thereof, has become unsustainable for both preservation and day-to-day management. This has stifled discovery of our digital collections and halted our ability to push technological boundaries for digital scholarship and develop creative projects with our data.

DAMP! anticipates that the creation and curation of digital content will grow exponentially over the coming years, as our institutions evaluate campus needs and determine the roles the Libraries can, and should, play. To that end, DAMP!’s digital asset management and preservation plan is informed by:

- The strong desire to make collections and their content findable in a more consistent way.
- The need for policies and best practices surrounding management and preservation.
- Issues surrounding the sustainability of our systems, most notably that the development and customization of any one system requires extensive programming support, expertise and time.
- Concerns that our storage needs cannot continue to grow so exponentially without long term, coordinated planning.
- Our relative lack of digital preservation capabilities across our current DAMS.

As we determine and prioritize the needs of our users and ourselves and the best ways to address them, DAMP! is cognizant of the range of potential issues associated with selecting, developing, and maintaining digital solutions. We recognize that solutions should be adopted based on either their capacity to enhance current research, teaching, and learning activities on campus, or their capacity to provide new and dynamic pedagogical methods. We know that technology is not the universal remedy. A digital asset management solution must consist of
both a technical infrastructure and a strong staff network of relevant expertise. It must be guided by principles and best practices.

A change in how the TriCo does digital asset management is needed in order to best serve our user communities, to keep up with our peers, to follow best practices, preserve valuable assets, and to become innovators in our own right. The Digital Asset Management and Preservation working group (DAMP!) recommends that the TriCo facilitate discovery and use, optimize staff time, and become forward thinking with regards to our collections by consolidating our DAMS into Hydra/Fedora.

We acknowledge the need for a comprehensive digital asset management strategy which creates the infrastructure in which effective practices can be implemented. There is now an opportunity for us to be fiscally responsible and proactive.

In the report that follows, DAMP! has taken a thorough look at the current landscape of digital asset management systems and preservation technologies, including our own environments. Our recommendation has been based on evaluation of best practices, training with leading experts in the field, careful analysis of systems, and discussions with our peer institutions. In addition to recommending Hydra/Fedora, we detail a digital asset management system strategy for the TriCo, offer recommendations for appropriate investments in tools and systems, and highlight necessary staffing and expertise.

Background

As early as 2000, the libraries, archives, and museums (LAM) communities began efforts to define the attributes of a trusted digital repository. Having long been entrusted with safeguarding the materials and objects that document our cultural heritage, the LAM community faced enormous challenges in attempting to reliably preserve and provide access to digital materials. In 2002, the Research Libraries Group (RLG) and OCLC published “Trusted Digital Repositories: Attributes and Responsibilities,” a report which articulated a framework of attributes and responsibilities for trusted, reliable, sustainable digital repositories. In the report, the authors explain that digital resources need more attention that resources on paper, and often much sooner. Their inherent fragility leaves a small window of opportunity for action to be taken before resources are lost.1

Thirteen years later, technologies have evolved, but many of the fundamental issues around preserving digital resources remain the same. Digital materials are being produced rapidly and in tremendous volumes, and remain at risk. For instance, the hardware and software needed to read digital materials can become obsolete very quickly, thus digital materials require an active approach to management and preservation. Both the framework developed by RLG-OCLC and

the OAIS model\(^2\) demonstrate the critical importance of a coordinated, comprehensive approach to digital preservation with the appropriate administrative, organizational, financial, and procedural resources in place to complement the appropriate hardware and software; a technological solution alone will not suffice.

To meet the growing demand from libraries for solutions to managing digital materials, numerous digital asset management systems (DAMS), both open-source and proprietary, were developed. These systems had varying capabilities, including providing metadata records, storage, searchability, web presentation, discoverability, preservation, analytics, as well as inventory management for special collections.

Unlike the integrated library systems used to manage traditional library materials, there existed no comparable comprehensive “out-of-the-box” system to manage unique or original collections. Different collections required different standards, and different departments had different needs. As a result, throughout the TriCo, we saw a preponderance of uncoordinated solutions to digital asset storage and dissemination being adopted. Through the years, the LWSI and TAG committees (among others) regularly authorized reports which questioned whether it made sense to use so many different systems to manage digital content, but necessity required that we maintained the approach.

By 2014, we were struggling with the basic management and discovery of our growing body of assets and the sheer number of software packages managing those assets. It is an unsustainable infrastructure. We can no longer ensure the long term survival of our born-digital and analog-to-digital objects without adding additional systems, nor can we scale any of our existing software to meet all of our TriCo needs, particularly those that serve faculty, students, and external user groups.

Consistent with this, the Digital Asset Management Assessment Group (DAM) prepared an overview of the DAMS in use in Fall 2014, and made a series of recommendations on the storage and management of the TriCo’s diverse and burgeoning digital assets. Their report additionally explored the wide variety of item types and digital formats in various systems both locally and externally hosted, and determined that while most of the current TriCo systems were capable of managing many of these material types, no one system was appropriate for all assets.

In January 2015, stakeholders from the Tri-Colleges attended a three-day workshop led by Nancy McGovern and Kari Smith. In response to this workshop and to the recommendations made by the DAM Group, the TriCo’s leadership charged a Digital Asset Management and Preservation Working Group (DAMP!). This group would develop effective practices for the management and preservation of digital assets stewarded by the libraries and campus partners. Furthermore, they would provide recommendations and a strategic plan to develop a simplified

---

and sustainable infrastructure (including software, budget, and policy framework) for our digital assets.

**Current Landscape: Limitations and Opportunities**

At present, we use four separate systems for digital asset management; two additional systems provide management for art and artifacts that have different metadata and database needs. **CONTENTdm** is used for digitized archives, manuscripts, and rare book materials; **DSpace** and **bepress** are used as institutional repositories and comprise faculty and student scholarship; **EmbARK** is used as an art and artifacts collections manager as well as a digital asset management system for digitized materials online; **ArtStor/Shared Shelf** is used for visual resources for classes as well as faculty projects that require an image database; **Portfolio** is used as a metadata development system and digital collections manager. For more detailed information on the systems we currently employ, please see the [Content Map](#).

**Cost of Current Systems**

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>bepress Digital Commons</td>
<td>~$16,000 / institution</td>
<td>~$16,450/year / institution</td>
<td>~$16,915/year / institution</td>
</tr>
<tr>
<td>CONTENTdm</td>
<td>$9,396</td>
<td>$10,025/year</td>
<td>$10,325/year</td>
</tr>
<tr>
<td>DSpace</td>
<td>$0</td>
<td>$21,375/year</td>
<td>$37,125/year</td>
</tr>
<tr>
<td>EmbARK Collections Manager / Web Kiosk</td>
<td>$8,350</td>
<td>$20,000/year</td>
<td>$22,000/year</td>
</tr>
<tr>
<td>Portfolio</td>
<td>$1,340</td>
<td>$19,340/year</td>
<td>$28,500/year</td>
</tr>
<tr>
<td>Shared Shelf</td>
<td>$18,258</td>
<td>$18,808/year</td>
<td>$19,370/year</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>~ $85,344</td>
<td>~ $138,898/year</td>
<td>~ $168,065/year</td>
</tr>
</tbody>
</table>

*It should be noted that these are inaccurate estimates because our current configuration. Siloed systems, duplication of assets, and general lack of policies and communication make projections very difficult. These figures also do not reflect the added staff costs involved in inefficiencies of time and scale.*
Discovery and Interconnectivity of Data

Researchers face tremendous challenges when trying to use and access our digital assets. We as librarians know that many digital assets exist, distributed across the different colleges, with varying data structures and storage architecture, and tied to a variety of software applications.

DAMP! has had discussions with various library stakeholders who have expressed concerns regarding the overabundance of individual silos of digital objects, poor metadata, the quickly growing number of digital collections, IP-access limitations, web development platforms, and flexible APIs. With the implementation of a single flexible DAMS, we can develop a system to our TriCo specifications - specifically, better enhanced discovery of our digital objects. Fewer repositories would reduce our users’ confusion over where a digital object can be found.

Opportunities: Hydra/Fedora based discovery projects elsewhere
1. The Digital Public Library of America (DPLA), Stanford University, and the DuraSpace Hydra-in-a-Box Initiative (http://duraspace.org/node/2527)
2. Northwestern University Repository (Repository | images) uses a Hydra application that provides a repository solution for discovery of and access to images for the public, faculty, students, and scholars. Patrons with a university id can download and manipulate images; those without can view images and download those that are defined as public: https://images.library.northwestern.edu/
3. Amherst College has a Fedora-based digital repository and is in the process of adopting Hydra: https://acdc.amherst.edu/

Consistent Metadata Standards and Best Practices

The main purposes of metadata are to facilitate the discovery of relevant information and to provide a means to ensure the long-term preservation of information about an institution's collections. In addition, metadata is used to document information about copyright and use restrictions and track technical information such as digitization information and data encoding. Metadata is frequently governed by community-developed and community-fostered standards and best practices in order to ensure quality, consistency, and interoperability. At present, some of our siloed systems impose consistent metadata standards and some do not. While one can argue that consistency can be imposed without an integrated system, the practical fact is that at present this is dependent on the quality of cataloger input, from both students and professional staff, rather than automated quality control.

Information on metadata standards:
2. https://psap.library.illinois.edu/format-id-guide/metadata
Preservation

None of the systems have built-in preservation infrastructure, although DSpace does offer a minor preservation component with automatic checksum generation. An integrated system is essential if we want valuable assets to survive long term. Many granting agencies will not consider support of digital projects that do not include a sustainable preservation environment.

Born-Digital Institutional Records, Records Management, and Research Data Management

Within the Tri-College communities, numerous digital assets are being created every day. These digital assets include research data, instructional materials, and administrative and business records, including records for which the TriCo have legal and archival responsibilities. Digital content has become critical for day-to-day operations, and it constitutes a growing portion of the historical and archival documentation of the TriCo and each college’s activities and accomplishments.

At each of the campuses there are a number of different departments and units that have a vested interest in the management and preservation of our digital assets. Faculty whose research comprises large datasets have minimal support and no data repository exists. Primary grant funders like the NEH and NSF require technical standards, particularly regarding data management and sustainability. Institutional Advancement, College Communications, Athletics, Alumni Relations, and Admissions are continually trying to engage and create a sense of connectedness between faculty, students, staff, alumni, prospective students, and various other constituent groups. College publications, in the form of newsletters, magazines, admission viewbooks, and informational pamphlets, college social media technologies, and basic college web-based communications rely heavily on the use of and access to digital assets. Utilization of digital images and media and digitized institutional resources has become an invaluable engagement strategy that will only increase in years to come as more records are being created digitally. It is critical that the proper management, preservation, and access to these digital assets is insured well after they are accessioned into each of the institution’s College Archives.

A reality that the Tri-College communities face is the need to meet administrative, fiscal, legal, and historical obligations through the systematic and consistent management of all records, regardless of medium or format, created and/or maintained by its employees in the course of the college’s academic and administrative business functions. The essential nature of the modern office, complete with hybrid paper/electronic systems, digital environments established to support manipulation and repurposing of data at the expense of recordkeeping, obsolescence of hardware and software, media decay, and other problems, makes it difficult for the archivists, records managers, and librarians to provide long term preservation of authentic electronic records and to maintain the accountability of the institutions and operations which those records document. Born-digital records require ongoing management in order for them to remain viable and retain authenticity in an archival context. We currently have no infrastructure to support born-digital archives.
This nature of recordkeeping leads institutions to create and maintain electronic records that they cannot automatically trust and depend on in the same way that institutions trust and depend on traditional paper records. In general, archivists, records managers, and librarians have difficulty preserving electronic records that fail to be (1) accessible or readable due to compatibility and obsolescence issues; (2) identifiable and retrievable due to lack of sufficient metadata; and (3) reliable in the accuracy of their content due to the ease of updating and altering records, either inadvertently or purposefully. In order to address these issues, we must develop systems, policies, and infrastructure that protects and technologically stabilize the content and context of our electronic records.

A first step toward the preservation of electronic records is proper management of our digital assets. As stated the management of our digital assets is currently being done in disparate siloed systems none of which provided digital preservation services, which leaves our institutions vulnerable to data and record loss.

Opportunities: Hydra/Fedora based Preservation Projects elsewhere

1. HydraDAM2 Project at Indiana University/WGBH (http://www.avalonmediasystem.org/blog-post/hydradam2)
2. Hydra/Fedora Repository at Yale (http://web.library.yale.edu/lit/projects/hydra) or see the powerpoint: http://web.library.yale.edu/sites/default/files/files/HydraPreservationRev(1).pptx

Scalability of Increasing Digital Assets and Digitization Initiatives

Digitization is crucial to access and outreach and a key way for outside researchers to find our collections online. Additionally, we need to be able to sustain the digitization we have done so we do not lose important work, including having to re-scan collections that were digitized because of their fragile conditions in the first place, and collections that are in high use. Digitization also contributes to rich digital scholarship projects and new forms of scholarly inquiry. We need to be able to adapt our infrastructure to be more efficient while remaining flexible. In the past 15 years, we have digitized over 10TB of materials and continuously pursue digitization grants and projects to provide more access to our collections. Despite our best attempts to predict storage needs, in 2008, 2010, 2013, and 2015 we have exceeded capacity on our DAMS and on our local servers. Our current storage infrastructure does not allow us to easily add additional storage. Our digital assets are only going to grow in number as well as size. This does not include out intended stewardship of born-digital records collections, which also grow at exponential rates.

Ability to integrate new formats with existing collections

We are at a standstill around archiving hi-def video associated with some projects like Minding Swarthmore (a film produced in conjunction with Swarthmore’s Sesquicentennial) because none of our systems are quite adequate in terms of storage, description and access. Currently, we have to choose whether to store materials adequately, apply appropriate, consistent metadata or make them accessible.
Digital Scholarship

Digital Scholarship advances research by incorporating into traditional scholarly inquiry emerging digital tools and processes such as data visualization, network analysis, text-analytic techniques, GIS/mapping, and data-mining. As these new and important curricular, co-curricular, and research projects emerge and as multimodal instruction and learning is becoming more attractive to faculty and students there is a greater demand for the use, access, and repurposing of our colleges' digital assets, which open up rich possibilities for cutting-edge research by faculty and students. We strive to support the uses of these tools and techniques to help faculty in their approaches to research and classroom pedagogy, and to support students in their development at researchers, critical thinkers, and knowledge creators.

Our current digital asset management systems place a limit on those possibilities. Currently, our digital scholarship librarians spend an enormous amount of time developing technical expertise to find ways to push and pull digital assets from our disparate systems. Because our digital assets are currently held in a number of different systems, access to the data produced from digital scholarship project is severely limited. The TriCo wants the work that students and faculty are creating to be available for use, and there is a growing demand for the back data that has become a scholarly resource not only for the everyday researcher, but also for large digital asset aggregators like the Digital Public Library of America.

While we have always been on the leading edge of collaborative digital scholarship projects, we are slowly giving up that edge because we are technologically limited by the systems we use to make our data more interconnected. The growing digital scholarship community is embracing new technologies and creating new ways of presenting data to digital humanists, which we cannot take advantage of due to system, technology, and expertise limitations. If we want to be seen as leaders in digital scholarship, our systems must facilitate the variety of digital scholarship projects now and in the future.

Recent or ongoing Tri-College digital projects provide some illustration. All of these involve extensive student and faculty work and have received national attention. However most exist as individual exhibits and the digital assets cannot easily be re-used:

Black Liberation 1969 Archive
http://blacklib1969.swarthmore.edu/
This site is a digital archive designed in support of Black Liberation 1969: Black Studies in History Theory and Praxis taught at Swarthmore College by Professor Allison Dorsey. Nabil Kashyap (Librarian for Digital Initiatives and Scholarship) built the site using open source software packages Omeka, TimelineJS, and Neatline. Students scanned hundreds of
photographs, newspapers, etc. but those images exist only in the Omeka “archive” and cannot be discovered separately in any TriCo discovery system. This site is not sustainable long term. And because these digitized materials were not managed, some items were scanned three times resulting in clear duplication of effort. Any reuse of the materials will require rescanning.

Ticha Project

http://ticha.haverford.edu
This tool will make Colonial Zapotec texts accessible to scholars in diverse fields (including linguistics, anthropology, and history), Zapotec community members, and the general public. Ticha will eventually allow a user to access and explore many interlinked layers of these texts, including images of the original documents, transcriptions, translations into English and modern Spanish, linguistic analysis (including morphological interlinearization), and commentary. Ticha is innovative in bringing together data analyzed in FLEx (Fieldworks Language Explorer, fieldworks.sil.org) a system for lexical and grammatical analysis, with current TEI standards (Text Encoding Initiative, tei-c.org) for paleographic and translational representations of texts. This is a stand-alone site using materials that are not owned by Haverford and is not integrated into the TriCo discovery systems.

College Women

www.collegewomen.org
College Women, currently in the pilot phase, is a consortial project of the libraries and archives of the institutions formerly known as the Seven Sisters and led by Bryn Mawr College. The project began in the spring of 2014 with a Foundations Grant from the National Endowment for the Humanities to support the development of an online portal capable of searching the institutions’ digital collections of student material via their local repositories’ APIs. The intention of the site is to open new avenues for research in American women’s history by making the dispersed writings, images and documents of women students easily accessible through a single search. When brought together, the collections will enable new studies in political reform and women’s rights, sexuality and body image, religion, race and class, as well as major domestic and international events. These documents form the core of a collection that will grow, evolve, and become an essential resource for scholarship on the impact of education on the trajectory of women’s lives.

Quakers and Native Americans, 1791-1815

http://triptych.brynmawr.edu/cdm/landingpage/collection/SC_Natives
Still in its formative stage, this project will make available hundreds of original journals, correspondence, minutes, and other contemporary records from both Haverford and Swarthmore Special Libraries that document extensive interaction between members of the Society of Friends and Native Americans in the late 18th and early 19th centuries. These materials have been digitized, transcribed and encoded in TEI in anticipation of a major conference on the subject that will occur in Philadelphia in November of 2016. The encoding will allow users to interact with the materials in time, space, and between individuals and groups. Students have been involved from the beginning in the transcription and TEI markup. Scholars in the US and abroad have shown great interest in the material. We currently have no
In summary, humanities and digital librarians have done a good job with what is available, but are frustrated with the limitations of the current configuration and cite a need for:

- “Faster delivery of digital archives and digital exhibit based projects.” Right now these types of projects involve reinventing digitization, metadata and storage procedures each time, which is time consuming, has a high barrier to entry and often involves duplicating effort. A more flexible system that met current standards would save time and allow more faculty and classes to make use of this genre of digital project.
- “Sustainable digital projects.” Right now digital archives and digital exhibit based projects are spun off individually and are in no way connected to our actual repository. With Black Liberation site referenced above, for example, there are no archival digital surrogates because there was no space or capacity to generate them in the first place.

Decision Criteria

In order to compare DAM systems “apples-to-apples”, we evaluated each system based on a list of functional requirements. The functional requirements are divided into system supports, system allows, and preferred features.

System supports

1.) Import/Ingest
   a.) Ingest content files easily via a web interface
   b.) Specify the content is for:
      i.) preservation and publication
      ii.) dark storage/preservation only
   c.) Persistently manage and store digital object including, but not limited to, object with images, audio, video, and text content file formats.
   d.) Upload preservation/master copies of content files (e.g. TIFFs) and have derivative/service copies automatically generated when possible (e.g. JPEG or other thumbnail images) instead of having to create separate service copies.
   e.) Upload individual content files -- or batch upload multiple content files -- with minimal to almost no metadata, and subsequently come back and add more robust metadata to the objects.
   f.) Metadata embedded in content files to be extracted on import into the system, for ease of working with and adding to resulting metadata record.

2.) Metadata and Object Modeling
   a.) The import of legacy database records and metadata crosswalks
   b.) Macro automation (e.g. export, save, localize) customizable
   c.) Upload individual content files-- or batch upload multiple content files--and simultaneously upload and associate metadata records.
   d.) Import of legacy/local authority records and maintain authority control
e.) The ability to batch import and export bitstreams and metadata into the DAMS
f.) Ability to link digital objects to finding aids and other relevant and extant databases to DAMS through the metadata
g.) Ability to have mappable metadata schemas
h.) Ability to export metadata in standard formats for later reuse
i.) Ability to manage access restrictions at the item-level and collection-level as well as an embargo
j.) The metadata scheme will include, but not be limited to, data elements including: descriptive, technical, and rights information
k.) Ability to select and describe content from across multiple collections, creating artificial, topical collections within the local institution, and across the TriCo
l.) Metadata export tools for administrators to extract all metadata

3.) Publication and Access Control
   a.) Allow for publish and unpublish of a digital object; unpublished digital objects would only be available for viewing by curators and staff using the system to create and manage object before making them publically available.
   b.) Publishing of digital objects with a quick turnaround time for availability to end users, to be responsive for faculty request for objects, users with scan-on request with a mechanism to handle requests for reproductions/licensed use.
   c.) Control the degree of user access to a published digital object, based on the access level designated in the metadata record for the digital object.
   d.) Limit what is shown to end users at different access levels; for example, show only the metadata record for a given digital object--but not the content file(s)
   e.) Allow for changes in access controls for digital objects.
   f.) Published digital objects to display full metadata records, so users have complete descriptive, rights, etc. information for an informed use of the resource.
   g.) A mechanism that monitors embargoed digital objects

4.) System Administration and Management
   a.) Define user accounts with different permission levels (e.g., full admin privileges, metadata editing only, etc.), for different staff that will be creating digital objects (e.g. student assistants)
   b.) Ample digital asset storage space for each institution
   c.) Access to user statistics (e.g., number of hits, terms queried, location)
   d.) Access to campus repository statistics
   e.) Ability to assign different parts of workflow to internal experts (e.g. metadata to a metadata librarian)
   f.) Ability to track hours spent on metadata creation, preservation, and other workflows.

5.) Search and Discovery
   a.) Easy navigation to different sections/collections
   b.) Basic and advanced search: narrow by various fields and faceted searching/browsing. All metadata fields exposed in the search results should be
available to be searched independently or in combination from an advanced search page. The exposed metadata fields will include all valuable field in a given metadata schema.
c.) By default searches should be conducted across all collections with the option of limiting to a specific collection.
d.) Users can download object bit streams into local applications and systems, within limitable rights.
e.) Multi-staff and departments’ involvement in workflows.
f.) Ability to have full-text search and full-indexing with prioritized search terms.
g.) OAI-PMH compliant. Harvest digital objects using an OAI interface in order to gather all of the metadata.
h.) Robust API.
i.) Options for content delivery based on required access levels (thumbnails, downloads, no copying, etc.)—onsite only?
j.) Section 508 compliance (accessibility).

6.) Preservation
   a.) Preservation services and infrastructure such as: checksums, checksum validator, PREMIS, virus checks.
b.) BagIt, technical metadata extraction.
c.) Permanent URLs.
d.) A clear data recovery process.
e.) A variety of digital objects including: images, audio, video, text, PDF, WARC, compound.
f.) Authentication-IP based, LDAP, and Shibboleth.
g.) OAIS compliance.
h.) Ability to become Trusted Digital Repository (TDR).

System allows
   • The ability to have an internal link between the high-resolution file and the access file.
   • System has active and enthusiastic user community.
   • Site responsiveness.
   • Editing, validation, and display of TEI text.
   • Related item discovery.
   • Syndication for end-users/RSS Feed.
   • RDF (standard for encoding metadata and other knowledge on the Semantic Web or “linked data”).
   • Integrated use of authoritative vocabulary terms/headings from Library of Congress, Getty Visual Resources Association, Thesaurus of Geographic Names, or locally created headings within metadata records.
   • Page turner feature for complete works such as books, manuscripts, scrapbooks, etc.

Preferred features
   • Complex datastream management to show various relationships between objects and collections.
- Integration of donor dropbox for born-digital content
- Ability to create rich forms for XML schemas, including multiple forms of specific schemas for different users or viewable in a single form
- Ability to create temporary sets of files for cataloging
- Side-by-side item viewing
- Save items to user workspace (user generated galleries)
- Transcription widget
- Share this widget

Systems Considered

Modification of Existing TriCo Systems

DAMP! investigated whether we might be able to make modifications to our existing software to meet our current and future needs (such as interoperability or digital preservation.) We looked to connect our current systems either by APIs or data harvesting. Not all of the systems have APIs and the ones that do are fairly inflexible. We examined our six major systems: CONTENTdm, Portfolio, bepress, DSpace, EmbARK, and SharedShelf. CONTENTdm is the only system which has its materials harvested through OAI-PMH into Tripod. Its API can only output Dublin Core in the way that it is set up in the system and cannot be altered to match a new schema. Compound objects with varying metadata based on a page, for example, do not export at all. DSpace is a very interoperable system but its search functionality is not user friendly. Portfolio, bepress, and EmbARK do not have APIs and would have to remain siloed. DSpace and bepress have significant overlap in functionality and purpose as well. Additionally, DSpace is the only system with preservation capabilities such as checksum generation and all preservation workflows will need to be done around the other systems.

Hydra/Fedora

Hydra is a technical framework built for the development of digital repositories, and is a platform upon which our DAMS could be built. The Hydra project addresses the idea that no single system can provide the full range of repository-based solutions for a given institution's needs, so it offers the ability to flexibly satisfy the demands of any collection. Hydra offers several software libraries written in the Ruby programming language, which is especially known for allowing rapid development. According to Tom Cramer, the founder of Hydra and the Chief Technology Strategist of Stanford University Libraries, Stanford went from one to eight Ruby savvy developers in one year with no new hires, and estimates a one week learning curve to basic proficiency. There are many diverse examples of projects that can be built with Hydra that are found here. University of Virginia's Libra, their online archive of scholarship, demonstrates how their platform used Hydra to organize a large amount and wide variety of scholarly output. Members of their community are able to explore content with a faceted search and submit Open Access (OA) works, electronic theses and dissertations, and datasets. The Northwestern University Image Repository is a project with rich content and useful features, such as the ability...
to build custom collections of images, and export a custom collection to a Powerpoint. Hydra is open source, allowing any developer committed to the project to contribute. It is a cross-institutional effort, and has been developed to satisfy the numerous requirements of many diverse digital collections. Its highly active and dedicated community of developers and adopters are constantly extending and enhancing its core. Hydra seems to be a better organization than the alternatives we have considered.

Several needed aspects of a digital repository, such as the user interface, rich metadata editing, authority control, automatic generation of derivative copies, and the management of object relationships, are not part of Hydra. All of this would require extra development, and could be satisfied by a website plugin called Blacklight. Blacklight can be utilized by Hydra to provide an interface for users to discover and browse digital content. It is able to accommodate many different formats of digital objects and display them appropriately. Blacklight offers a simplified browsing experience for users, allowing them to easily search for data with advanced queries and save content with stable URLs for both search and record pages. Several examples of the plugin in use within a variety of projects can be found at the Blacklight Project Showcase. The United States Holocaust Memorial Museum Collections Search shows how Blacklight helps manage their massive amount of digital content in a wide variety of formats, with several categorical options for finding specific oral history recordings, films, documents, names sources, and many other formats. The North Carolina State University Libraries' Rare and Unique Digital Collections is efficiently organized such that a researcher can find relevant historical information of North Carolina pertaining to a specific decade, person, city, street, and more. The rich user interface possible with Blacklight could satisfy all the discovery and display requirements identified by DAMP! (POT 1 LT 1C, 2013) with some development effort.

Hydra relies on Fedora (Flexible Extensible Digital Object Repository Architecture) as a robust layer for managing digital objects. It is widely used by university libraries, national libraries, research institutions, government agencies, and cultural heritage organizations. It enables users to access large and complex collections of historic and cultural digital materials. It can act as a preservation system with configurable security architecture and options to connect to possible long term storage and archiving methods in the future. Data validation and fixity can be developed within Fedora as well, providing more support for long-term archiving than what is currently available with our server setup. With Fedora, we would be able to implement linked open data and expose our data more widely. Linked open data is a method of publishing structured data so that it can be interlinked and become more useful through semantic queries. Like Hydra, Fedora is also open source and has a community of dedicated contributors. In relation to Hydra, Fedora can be best described as the underlying architecture for a digital repository. The complete management, indexing, discovery, and delivery of digital content are handled by the other technologies described here. The current version, which we would use, is Fedora 4.

Sufia would be an ideal Hydra add-on that allows for a multitude of useful features to be added to a digital repository. A few of these features include multiple file uploads, flexible user and group based access controls, and social media interaction. The project page has a complete list
of available features. These additional features would be ideal for creating any digital repository for an institution.

The Hydra/Fedora framework will also address the growing need for platforms on which faculty and students can create digital scholarship projects. One notable project that fosters academic collaboration and exhibition is the Penn State ScholarSphere. This platform enables students, faculty, and staff to post their academic papers, presentations, posters, and other scholarly creations to one location. With a discovery interface powered by Blacklight, users are able to easily explore and discover posted content by tags, subject, creator, and other categories.

Peer institutions such as Amherst College and soon Lafayette College are partners of the Hydra and Fedora projects, as well as other local institutions such as the Chemical Heritage Foundation, the American Philosophical Society, and Temple University. A complete list of partners of these projects can be found here. We spoke with these institutions and they were extremely enthusiastic of and supportive of these frameworks, as is the Hydra developer community more broadly. Also highly supportive are the Hydra and Fedora communities, who direct resources toward strategic projects or community needs, an approach which ensures that new features or applications are responsive to the communities. We already have experience using open source software with VuFind, an open source library search engine, and have not encountered any major problems.

**Islandora**

Islandora is an open source digital repository system primarily based on Fedora Commons and Drupal. It was developed at the University of Prince Edward Island and the major vendor for support, consultation, and development is discoverygarden (http://www.discoverygarden.ca/). It is for the most part a turn-key DAMS. Islandora’s stack is closed and does not allow for much modularity. Islandora's front end is based on Drupal, which has many modules but Islandora does not use Drupal nodes, so most Drupal modules will not work without significant programming. The back-end is Fedora and is treated like a module. Solution Packs are offered as add-ons for different content types. Islandora’s Tuque API relies on Fedora 3 code that will need to be completely rewritten for Fedora 4.

Peer institutions we spoke with also noticed a disparity between module development within the Drupal community and the Islandora community, particularly in regards to the upgrades to Fedora 4 and Drupal 8. Consequently, there are a limited number of modules that will work with both until the rewrites for Fedora 4. Overall, Islandora would be a good replacement for CONTENTdm alone, but is not extensible enough to have the multiple interfaces. For example, it suits Barnard’s needs because they only needed a repository for digitized college archives and do very few digital humanities projects beyond basic online exhibitions.

**VuDL**
VuDL (vudl.org) is an open source digital repository software developed at Villanova. Similar to Islandora, it is based on Fedora Commons and VuFind. We currently use VuFind as our discovery layer in Tripod, which was also developed at Villanova. The core of VuDL’s application is powered by Orbeon Forms, a powerful XML/XForms processor. eXist, a native XML database, and the server's file system combine to support the data and image repository. There is currently very little documentation or evidence of a strong user community. As of 2011, Villanova was seeking development partners (Library Technology Guides, 2011).

Recommendations

After careful consideration of the TriCo’s options (see Systems Considered), DAMP! recommends that we pursue an open source system with a stable, robust database for metadata and digital preservation coupled with the ability to create and modify any number of front-end interfaces for public discoverability and display. To this end, the Hydra/Fedora technology stack is the most robust and sustainable system for both our current needs and future growth opportunities through its “one body and many heads” approach. Combining the majority of our digital assets into one DAMS based on Hydra/Fedora will allow us to develop a holistic approach to digital preservation with a robust, configurable back-end based on collection needs. The Hydra/Fedora framework offers a way to combine management and storage efficiently in a more streamlined using a stable Fedora database with several customized Hydra heads. These heads are best adapted to serve the ever-changing needs of our college communities.

Given that the TriCo digital collections are growing at an exponential rate, DAMP! recommends strategic planning for our changing storage needs. A systematic review would include a decision between local server and cloud-based options as well as the possibility of implementing a layered storage solution, i.e. a “dark archive” for preservation purposes. Similarly, System Administration staffing levels should be considered and adjusted according to these decisions.

Furthermore, DAMP! recommends the development and adoption of formal digital preservation and asset management policies and workflows and the application of uniform metadata standards to support data maintenance and integration throughout the various custodial units of the TriCo. In order to fulfill this, we would form and build a digital preservation team, raise community awareness of our preservation initiatives and their objectives, and align our policies with archival best practices. The results would include a forward-thinking plan regarding digitization practices and accessioning born-digital materials. For more information on policy, workflow, standards development, and general planning see the Digital Asset Management and Preservation Framework Tasks document.

The completion of this new Hydra/Fedora DAMS will allow the TriCo to sunset several current DAMS which no longer support either the back-end functionality or front-end discoverability that top-tier institutions require in order to best serve our campus communities and meet best

---

practice standards. These database systems include CONTENTdm, DSpace, Portfolio, and bepress.

Although DAMP!’s recommendations rely on the integration of several digital object silos into a single repository, we do not feel that it is advisable at this time to merge all existing TriCo databases. For example, we do not intend to consolidate Tripod, our Integrated Library System (ILS) into the new DAMS. Whereas we appreciate that our patrons want a more Google-like search across all of our holdings both digital and printed, the hurdles of programming such a system given the differences in back-end metadata and front-end display issues would significantly increase both the cost and length of this project. A scheduled review of our ILS and the current landscape of library systems in 2017 will allow us another opportunity to gauge the feasibility of creating an integrated Hydra head for Tripod.

Four additional systems will not be integrated due to their functionality. For example we do not intend to encompass EmbARK’s collections or functionality into this initial Hydra/Fedora installation. EmbARK provides collection management for our art and artifact collections and has different metadata and database needs. Similarly, ArtStor and Shared Shelf are strong tools used for teaching with digital objects and giving users access to collections outside of the TriCo, and we will continue to produce digital exhibits and support digital scholarship through Omeka. These systems may be considered for integration at the end of the initial five year project plan.

A consolidated, extensible, and expandable system is required to fit the TriCo’s needs. If we do not move in this direction, we lose the opportunity to provide a stable and responsible environment for our institution’s records. The success of the new DAMS is dependent upon the establishment of the appropriate technical infrastructure, staffing, and relevant policies for a discoverable system to serve our campus communities. With planning, forethought, a supportive community, and adequate resources we can invest in a strong system to provide clear discovery and preservation.

Project Overview

The goal of the project is to build a robust, consolidated digital asset management and preservation system, coupled with improved metadata, standards, and policies.

The project would include:

Policy Development

- Creation and approval of a uniform TriCo metadata schema using Metadata Object Description Standard (MODS) and mapped to Dublin Core
- Digital preservation policy
- Digitization workflow and baseline specifications for digital objects
- Appraisal and accession policy for digital content/digital collection development
● Approval workflow for new digital projects with significant needs, e.g. one needing a substantial increase in storage space or personnel time

Systems Development

● Development of the Fedora and Hydra architecture using project programmers and a consulting firm (Data Curation Experts)
  ○ Functional requirements development
  ○ Prototyping
  ○ Usability testing
● Reassessment of storage capacity, infrastructure, and System Administration staffing

Collection Migration

● Digital object and collection metadata conversion and significant clean-up
● Deduplication of digital objects
● Batch migration of collections
● Curatorial evaluation of digital object collections which no longer meet the new TriCo digital accession standard

Systematic Phase Out and Launch

● Software training for content managers, SuperUser, systems administrators, etc.
● Phasing out of existing software packages
  ○ bepress
  ○ CONTENTdm
  ○ DSpace
  ○ Portfolio

In addition to software customization, the TriCo intends to “right size” our system. This would include evaluating our current storage infrastructure to determine if the TriCo would be better served by hosting some or all of its digital content in the cloud, e.g. with Amazon EBS services; hiring an appropriate number of developers to maintain and augment the new DAMS and discovery tools; and seeking to capitalize on current college work toward federated authentication, e.g. InCommon.

While Hydra/Fedora is a significant investment and implementation would be no small feat, considerable savings will be produced by creating a database that is consolidated, sustainable, and extensible. The sunsetting of several existing DAMS, both in their upfront annual costs as well as savings from staff needing to manage, as well as learn to manage, both the front and back-end of six systems will allow us to better spend our time and budgets.
Strategy

Given that a new installation of Hydra/Fedora will give the TriCo a digital repository that is tabula rasa, we intend to take this opportunity to curatorially evaluate our collections.

During the time necessary to hire and train project staff and begin development of a DAM system, we intend to gather a small group to modify and approve a single MODS metadata schema mapped to Dublin Core. Similarly a group will approve a digitization policy as well as a digital collection policy which will govern the quality of current and future digital objects. The TriCo will use these standards as we clean and migrate existing objects and collections and as we move forward with adding new objects.

Given the general state of our digital collections (multiple iterations of the same object within one or more DAM systems as well as servers, objects without images in CONTENTdm, incorrect or absent metadata, etc.) and the fact that we will already be scanning the collections as we migrate, we intend to take this time to clean as well. We anticipate that this will add a substantial amount of staff time to the migration, but as we stated above, this can run concurrently with the hiring process and initial Hydra/Fedora programming phase.

Rather than remove all of the digital objects/collections from one DAM, sunset the system, then begin the next migration, we will base our migration on digital object “types”. The advantage to this type of migration means that we will have the time and opportunity to carefully massage each collection - deduping, rearranging, creating accurate metadata, etc. as needed before migration. Given that the content in CONTENTdm, DSpace, Portfolio, and bepress overlaps, duplicates, and is in general only marginally clean, this steady approach will clean up decades of disorder, benefiting our staff as well as the public using the collections. Using this approach, we recognize that we will need to maintain all of the current systems until migration is complete. However, at the end of this systematic approach, there should be no “orphaned” digital objects left in these current systems, allowing them to be sunsetted gracefully.

As we migrate our large collections from CONTENTdm, DSpace, etc., we recognize that there will be collections and items which don’t meet the standards for inclusion in our digital repository. The migration team and the library “owners” of these objects will agree upon a reasonable disposition of these objects.

We expect to complete the implementation stage in three years. In the project’s third year, we will transition to operational status. There will be a business model in place to support digital collections management. Employee time will be decreased after implementation is complete. However we recognize that following the full implementation, it may take one to two additional years to completely decommission the legacy systems.

The project team will do everything possible to mitigate these costs including seeking grant opportunities and partnerships with others in the Hydra Project community. There are currently
a number of similar grant-funded collaborative projects in progress, at institutions ranging in size from the Chemical Heritage Foundation to the University of Michigan.
Appendices

- Project Overview Timeline
- Literature Review
- Environmental Scan of Peer Institutions
- An Abridged History of Digital Assets at the Tri-College Libraries
- Content Map
- Preservation Assessment of Current Systems
- Generic Workflow
- Glossary
Project Overview Timeline
<table>
<thead>
<tr>
<th>Stages</th>
<th>Resources</th>
<th>Duration</th>
<th>Year 1</th>
<th>Year 2</th>
<th>Year 3</th>
<th>Year 4</th>
<th>Year 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Create TriCo Metadata Standard</td>
<td>2 TriCo committee members</td>
<td>12 months</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Set up Fedora and Internal and External Hydra Heads</td>
<td>2 FTE developers</td>
<td>24 months</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Clean up BMC collections (metadata and object deduping)</td>
<td>2 BMC FTE content specialists</td>
<td>30 months</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Clean up HC collections (metadata and object deduping)</td>
<td>2 HC FTE content specialists</td>
<td>30 months</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Clean up SC collections (metadata and object deduping)</td>
<td>2 SC FTE content specialists</td>
<td>30 months</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Develop TriCo Digital Preservation Policy</td>
<td>TBD</td>
<td>6 months</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Develop TriCo Digitization Standards</td>
<td>3 TriCo committee members</td>
<td>6 months</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Formalize Digital Object Collection Policies</td>
<td>3 TriCo committee members</td>
<td>12 months</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Formalize Digital Object Migration Workflows</td>
<td>3 TriCo committee members</td>
<td>6 months</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Formalize Digital Object Production Workflows</td>
<td>3 TriCo committee members</td>
<td>18 months</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Iterative Testing</td>
<td>TBD</td>
<td>18 months</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Deployment</td>
<td>2 FTE developers</td>
<td>12 months</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Storage Set-Up</td>
<td>1 FTE systems administrator</td>
<td>12 months</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BMC Collection Migration</td>
<td>1 BMC FTE content specialists</td>
<td>18 months</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HC Collection Migration</td>
<td>1 HC FTE content specialists</td>
<td>18 months</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SC Collection Migration</td>
<td>1 SC FTE content specialists</td>
<td>18 months</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sunset redundant systems: CONTENTdm, DSpace, Portfolio, bepress</td>
<td>1 FTE SuperUser; 1 FTE systems administrator</td>
<td>6 months</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Maintenance and Future Development</td>
<td>1 FTE developer; .5 systems administrator</td>
<td>Ongoing</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Literature Review


This is an exceptional article on the AIMS Framework, a methodology for stewarding born-digital materials. The project itself was inter-institutional and international – involving archivists, digital archivists, technical developers and repository managers, and other stakeholders within the University of Virginia, Stanford, Yale, and the University of Hull. The AIMS Framework is divided into four main functions that constitute sequential steps in a very high-level best practices born-digital workflow: collection development, accessioning, arrangement and description, and discovery and access. I found the first and last sections -- collection development & discovery and access -- the most interesting because they present possibilities and limitations that differ in many ways from those of analog records. There are also a number of useful Appendices, including sample processing workflow diagrams and an analysis of tools. These sections are supplemented with examples and discussion of related tools. Appendix F, for instance, includes a very good model for a Donor Survey and guidelines for creating donor agreements.


The purpose of Bergin’s sabbatical project was to identify institutions with established digital preservation programs and investigate how the programs were implemented. She conducted a web survey and 12 follow-up phone interviews. The survey found that only 25% of the institutions surveyed had a written digital preservation policy, but over 90% have taken some action to preserve digital materials. The most common strategy for digital preservation was replication, followed by bit preservation, migration, normalization, and emulation. The majority of institutions preserve digital materials locally, but some are participating in a combination of local and collaborative efforts. The most common level of staffing for digital preservation was 1 FTE. The most commonly offered digital preservation services were long term preservation of digital materials created by faculty, staff, students, and the user community, consulting on digital creation best practices, and preservation of institutional records.

Bergin’s phone interviews provided more in depth information on issues including funding, staffing, and digital preservation systems and tools. Most interviewees reported no direct funding for digital preservation and trying to fund it through existing budgets. Regarding staffing, most institutions were trying to manage digital preservation using existing staff that have other primary job responsibilities. Systems and tools being used included Archivematica, Fedora, LOCKSS, Islandora, Archive-It, Omeka, DuraCloud, Archivists Toolkit, PeDALS, MetaArchive, BagIt, and BitCurator.


The first edition of this report chronicled the evolution of preservation metadata from concept to standard, culminating with the release of the PREMIS Data Dictionary. Due to emergence of the PREMIS Data Dictionary as the de facto standard for preservation metadata, this second edition shifts focus from conceptual issues to implementation issues. The report provides an introduction to the Data Dictionary and notes that it must be tailored to meet the requirements of
a specific repository context - “it is not an off-the-shelf solution”. The most widely used framework for packaging preservation metadata is METS, an XML implementation of an OAIS information package. The PREMIS Editorial Committee has published guidelines for using PREMIS with METS. A number of tools exist which can extract metadata from digital objects and output PREMIS XML. Many tools, such as JHOVE and DROID, require post-processing (with other tools such as PREMIS Creation Tool or HandS) to convert output to PREMIS. The number of PREMIS implementations has increased in recent years. However, there is much variation in the types of digital objects for which PREMIS is used. In addition, the PREMIS functionality realized in each implementation varies considerably. Common usages include authentication using fixity information, validating the formats of digital objects, checking format migrations, provenance verification, and as a packaging mechanism for technical and administrative metadata. The report concludes by looking ahead to future work in preservation metadata and lists two areas in which increased attention would be beneficial: accumulation and consolidation of best practices and more information on the costs and benefits of preservation metadata.


This article stresses the importance of acting quickly with archiving born-digital materials and not waiting for the perfect solution. It provides beginning actions to take:

- Separate digital files from the media they arrived on: inventory # of bytes, implement storage (different instances for archival masters and access copies), document all actions
- Capture essential information about files by capturing checksums at point of transfer
- Create policies for acquisition of born-digital materials including preferred formats, consistent naming of files, organizational structure of materials and transfer medium


A Framework of Guidance for Building Good Digital Collections provides an overview of what is involved in creating and managing good digital collections, and identifies resources for the development of local practices for doing so.

The framework provides criteria for goodness in four areas: collections, objects, metadata, and initiatives (i.e. programs or projects to create and manage collections). It should be noted that “services” has been excluded; if the criteria is followed, all manner of services should be possible. For each of these areas, general principles are defined, and resources with further information are provided. In terms of collections, goodness involves curation, availability, and sustainability. Good objects are preservable and meaningful outside of their original context. Good metadata conforms to standards, is interoperable, and supports the preservation of objects. Adherence to best practices and consideration of its entire lifecycle are among the ways to define a good initiative.

This report traces the history, salient features, and impact of the OAIS reference model. The model was originally developed as part of efforts to develop formal standards for the long term storage of digital data generated from space missions and has since become accepted as the lingua franca of digital preservation. The report is careful to point out that OAIS is a model and not an implementation; therefore, the meaning of OAIS-compliance is vague. A key element of the model is its flexibility, although this seems to come at the price of consistency in implementation. Despite differences in implementations, the model has facilitated the development of standards, protocols, and best practices that form the foundation of an interoperable network of digital archives. Though the model was revised in 2012 from the original 2002 version, there were few significant changes. The report concludes by asking whether additional benefits would be obtained by establishing a more concrete relationship between the model and real-world implementations through standardization of the model’s concepts.


This paper recommends undertaking incremental actions immediately rather than waiting for a comprehensive solution for digital preservation and reviews tools for doing so. For institutions with financial resources and staffing Preservica and MetaArchive were recommended. Preservica is vendor solution for ingesting, accessing, and storing (using Amazon S3 + Glacier, though other back end storage can be used). MetaArchive is a community-owned private LOCKSS network (Stanford), which an institution joins for geographic distribution of multiple copies - the joining institution provides server space for other institutions’ materials in return for offsite storage of own materials. Being part of the community, which engages in active communication about digital preservation, is cited as more important than the storage.
An Abridged History of Digital Assets at the Tri-College Libraries

2000: With a grant from the Mellon Foundation for a range of cooperative activities, librarians begin discussing digital collections featuring materials from the Colleges' Special Collections.

2002: With funding from the SNAVE Foundation, Haverford begins a pilot project to digitize the Cope-Evans Family Papers. Triptych, our local installation of CONTENTdm, is born.

2002: The Tri-College Core Metadata Schema is established.

2003: Triptych is made public.

2004: The digitization of art images moves from Swarthmore's Art Department to the Libraries.

2005: TriDid, the first sustained experiment in managing a central repository of Tri-College pedagogical images, comes online using the open source system MDID. This efforts is seen as a natural step in uniting TriCollege digital image production and support. Not long after, the Libraries jointly license ARTstor.

2005: Using RLG's Archive-It service, the Libraries begin archiving Quaker and College websites.

2006: The Consortium agrees to serve as a test site for Variations3. The system is used to support the music programs at Haverford and Swarthmore, and a significant amount of on-site digitization of music and scores is undertaken.

2006: The Haverford Thesis server, established in 2003, needs to be replaced, and there is some interest in replacing it Tri-Collegially. The Tri-College Libraries Management Group agrees to support implementation of a test instance of DSpace. In the Fall, Haverford theses are added on a trial basis. During the Winter, all theses are migrated from EPrints to DSpace.

2007: TriDid is replaced by EmbARK, in an effort to provide a more robust cataloging interface for the digital images being produced locally. EmbARK Web Kiosk is added as an end-user interface, and it is released to the public that summer.

2007: It is recognized that the rapid increase in applications and services focused on storing and providing access to high volume, archival quality digital media will require significant changes to the storage infrastructure of our network. Our archival and functional storage requirements are expected to increase at least 5 terabytes annually beginning in FY ’10.

2007: The CONTENTdm Advisory Group is formed.

2008: CONTENTdm runs out of space. The Libraries question whether it makes sense to use three different systems to managing digital content. Because of our belief that that different
communities have different needs, we maintain this approach, but there is a recognized need to create of a formal assessment plan for digital archiving solutions and strategies.

2008: A digital image collection development and access policy is drafted.

2008: The possibility of a Tri-College Libraries' project focusing on art and artifact collections is discussed, as Bryn Mawr is seeking grant opportunities.

2009: The art and artifact collections project is revisited. The project could now build on six months of work done at Bryn Mawr and Haverford using EmbARK Collection Manager.

2010: We reach the 50,000 image limit in CONTENTdm and must pay an unanticipated $30,000 to continue to add items to the system. The Libraries question whether it makes sense to use three different systems to managing digital content. It is determined that migration is too challenging.

2010: A small group is charged to decide if we should license ARTstor's Shared Shelf.

2010: A group is formed to assess the types of collections and materials that are cataloged using non-MARC cataloging systems. Their final report includes suggestions for standards and best practices that will help us better create and manage metadata across the TriCo and the recommendation that we overcoming and accept the seemingly redundant nature of digital collections information.

2012: EmbARK Cataloger is phased out with the intention of replacement by Shared Shelf.

2013: DSpace and ContentDM run out of space.

2014: The Digital Asset Management Group is formed to revisit the 2010 non-MARC cataloging systems report.

2015: The Digital Asset Management and Preservation Group is charged.
Content Map
<table>
<thead>
<tr>
<th>SC_FHL</th>
<th>SC_SCPC</th>
<th>SC_McCabe</th>
<th>BC_Special</th>
<th>HV_Special</th>
<th>NOTES</th>
</tr>
</thead>
<tbody>
<tr>
<td>vendor, a few local copies</td>
<td>vendor and local copies</td>
<td>vendor</td>
<td>faculty scholarship, college archives, graduate dissertations, out-of-copyright German dissertations (19th century)</td>
<td>faculty publications</td>
<td>easy to use, appearance; strong SEO, 10tb storage included, peer pressure, has customizable metadata schema</td>
</tr>
<tr>
<td>CONTENTdm local drive, B: drive</td>
<td>CONTENTdm local drive, B: drive</td>
<td>CONTENTdm local drive, B: drive</td>
<td>CONTENTdm local drive, B: drive</td>
<td>CONTENTdm local drive, B: drive</td>
<td>page turning, compound object ingestion, transcription, viewer</td>
</tr>
<tr>
<td>pdf, txt</td>
<td>pdf, txt</td>
<td>pdf, txt</td>
<td>pdf, txt</td>
<td>pdf, txt</td>
<td></td>
</tr>
<tr>
<td>CONTENTdm</td>
<td>CONTENTdm</td>
<td>CONTENTdm</td>
<td>CONTENTdm</td>
<td>CONTENTdm</td>
<td></td>
</tr>
<tr>
<td>public discovery</td>
<td>internal DAM &amp; public discovery</td>
<td>College publications, student-run journals</td>
<td>College Archives publications (also in Bepress), College memorabilia, manuscripts collections (including letters and diaries), College Archives photograph (also in CONTENTdm)</td>
<td>College Archives, everything</td>
<td></td>
</tr>
<tr>
<td>PDF</td>
<td>PDF</td>
<td>PDF</td>
<td>PDF</td>
<td>PDF</td>
<td></td>
</tr>
<tr>
<td>Embark</td>
<td>Embark</td>
<td>Embark</td>
<td>Embark</td>
<td>Embark</td>
<td></td>
</tr>
<tr>
<td>jpg</td>
<td>jpg</td>
<td>jpg</td>
<td>jpg</td>
<td>jpg</td>
<td></td>
</tr>
<tr>
<td>West drawings &amp; Degas paintings to Triarte</td>
<td>fine and decorative arts, architecture, manuscripts</td>
<td>Fine and decorative art images</td>
<td>Needed a database to show art and artifacts, came with WebKiosk, metadata schema</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ArtStor, R: Drive</td>
<td>ArtStor, R: Drive</td>
<td>ArtStor, R: Drive</td>
<td>ArtStor, R: Drive</td>
<td>ArtStor, R: Drive</td>
<td></td>
</tr>
<tr>
<td>Shared Shelf</td>
<td>Shared Shelf</td>
<td>Shared Shelf</td>
<td>Shared Shelf</td>
<td>Shared Shelf</td>
<td></td>
</tr>
<tr>
<td>R: Drive</td>
<td>R: Drive</td>
<td>R: Drive</td>
<td>R: Drive</td>
<td>R: Drive</td>
<td></td>
</tr>
<tr>
<td>jpg</td>
<td>jpg</td>
<td>jpg</td>
<td>jpg</td>
<td>jpg</td>
<td></td>
</tr>
<tr>
<td>Art teaching images, faculty personal collections</td>
<td>Art teaching images, faculty personal collections, College Archives photographs</td>
<td>Art teaching images, faculty personal collections</td>
<td>Because we had ArtStor/beta offering, digital slide library</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Portfolio</td>
<td>Portfolio</td>
<td>Portfolio</td>
<td>Portfolio</td>
<td>Portfolio</td>
<td></td>
</tr>
<tr>
<td>R: Drive</td>
<td>R: Drive</td>
<td>R: Drive</td>
<td>R: Drive</td>
<td>R: Drive</td>
<td></td>
</tr>
<tr>
<td>PDF</td>
<td>PDF</td>
<td>PDF</td>
<td>PDF</td>
<td>PDF</td>
<td></td>
</tr>
<tr>
<td>Archive-it</td>
<td>Archive-it</td>
<td>Archive-it</td>
<td>Archive-it</td>
<td>Archive-it</td>
<td></td>
</tr>
<tr>
<td>capture web content</td>
<td>capture web content</td>
<td>capture web content</td>
<td>capture web content</td>
<td>wanted to archive the web effortlessly</td>
<td></td>
</tr>
<tr>
<td>Omeka</td>
<td>Omeka</td>
<td>Omeka</td>
<td>Omeka</td>
<td>Omeka</td>
<td></td>
</tr>
<tr>
<td>exhibits, digital scholarship</td>
<td>exhibits, digital scholarship</td>
<td>exhibits, digital scholarship</td>
<td>exhibits, digital scholarship</td>
<td>exhibited digital exhibits that could be easily produced by students</td>
<td></td>
</tr>
<tr>
<td>Archivist's Toolkit</td>
<td>Archivist's Toolkit</td>
<td>Archivist's Toolkit</td>
<td>Archivist's Toolkit</td>
<td>Archivist's Toolkit</td>
<td></td>
</tr>
<tr>
<td>accessions module</td>
<td>accessions module and finding aid creation</td>
<td>accessions module and finding aid creation</td>
<td>wanted to publish finding aids/PACSCL</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Preservation Assessment of Current Systems
<table>
<thead>
<tr>
<th>TRICO DIGITAL PRESERVATION ASSESSMENT</th>
<th>bepress</th>
<th>CONTENTdm</th>
<th>DSpace</th>
<th>Shared Shelf</th>
<th>Embark</th>
<th>Portfolio</th>
<th>Archive-it</th>
<th>Omeka</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>INGEST</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Copy [1]</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fixity Check [2]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Virus Scan [3]</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Auto Unique ID [5]</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>PROCESSING</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Auto Metadata Creation [6]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Auto Metadata Harvest [7]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rights Management [9]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Package Metadata [10]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Auto SIP Creation [11]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>ACCESS</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Public Interface [12]</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Auto DIP Creation [13]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>STORAGE</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Auto AIP Creation [14]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reliable, Long-term Bit Preservation [15]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Redundancy [16]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Geographically Dispersed Data Storage M [17]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Exit Strategy [18]</td>
<td></td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>MAINTENANCE</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Migration [19]</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Monitoring [20]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Auto Recovery [21]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TRICO DIGITAL WORKFLOW</td>
<td>bepress</td>
<td>CONTENTdm</td>
<td>DSpace</td>
<td>Shared Shelf</td>
<td>Embark*</td>
<td>Portfolio</td>
<td>Archive-it</td>
<td>Omeka</td>
</tr>
<tr>
<td>------------------------</td>
<td>---------</td>
<td>-----------</td>
<td>---------</td>
<td>--------------</td>
<td>---------</td>
<td>-----------</td>
<td>------------</td>
<td>-------</td>
</tr>
<tr>
<td>PRE-INGEST</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Appraisal and selection by staff</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Digital Assets Submission Agreement</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fixity (checksum)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Create a disk image to preserve order, contents and integrity or Write Lock directory of original files</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Transfer files to Ingest</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>INGEST - Create or Receive</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Automatically assign unique identifiers, error check (including fixity/checksum), validate files, virus check</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Assign unique identifiers, error check, and validate files by hand</td>
<td>X</td>
<td>X</td>
<td></td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Duplicate files with archival copy on R: drive</td>
<td>X</td>
<td></td>
<td></td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Point to archival copy on R: drive</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Automatically extract technical and tagged metadata</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Manually assign technical and tagged metadata</td>
<td>X</td>
<td>X</td>
<td></td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Manually assign rights metadata and subject headings</td>
<td>X</td>
<td>X</td>
<td></td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Create finding aid and bib record</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ARCHIVAL STORAGE</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Migrate files to preservation formats</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Transfer/copy files to Archival Storage (R: drive)</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Manage storage hierarchy</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Replace media as necessary</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Create access derivatives</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DATABASE MANAGEMENT</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Perform queries</td>
<td>X</td>
<td>X</td>
<td></td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Generate reports</td>
<td>X</td>
<td>X</td>
<td></td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Administer database (SuperUser)</td>
<td>X</td>
<td>X</td>
<td></td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Apply database updates (TriCo Systems)</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Apply database updates (Software supplier)</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Automated replication and integrity checking</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ACCESS &amp; DISCOVERY</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Manage public interface, web design (SuperUser)</td>
<td>X</td>
<td>X</td>
<td></td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Manage public interface, web design (TriCo systems)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Manage public interface, web design (Contractor or software supplier)</td>
<td>X</td>
<td></td>
<td></td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rights Management</td>
<td>X</td>
<td>X</td>
<td></td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EXIT STRATEGY</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Digital files and associated metadata are readily available to export (TriCo staff)</td>
<td>X</td>
<td>X</td>
<td></td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Digital files and associated metadata must be exported by software supplier</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* Used very differently on each campus
Glossary

**Archive**: To transfer records from the individual or office of creation to a repository authorized to appraise, preserve, and provide access to those records.

**Authorization**: The function of specifying access rights to resources related to information security and computer security in general and to access control in particular.

**Born-digital**: Information created in electronic format. A digital object that has never had an analog form. They differ from documents, movies and photographs that may have been scanned or converted to a digital format.

**Checksum**: An algorithmically-computed numeric value for a file or a set of files used to validate the state and content of the file for the purpose of detecting accidental errors that may have been introduced during its transmission or storage. The integrity of the data can be checked at any later time by recomputing the checksum and comparing it with the stored one. If the checksums match, the data was almost certainly not altered.

**Dark archive**: An archive that does not grant public access and only preserves the information it contains. This can refer to a digital archive or repository as well as brick & mortar archive.

**Digital asset management system (DAMS)**: The management, organization and distribution of digital assets from a central repository. It is a system that allows you to manage all your digital assets from one place.

**Digital object**: A representation of some piece of information in digital form. This can include many types of information, including word processing files, images, and digital audio files.

**Digital preservation**: The series of managed activities, policies, strategies and actions to ensure the accurate rendering of digital content for as long as necessary, regardless of the challenges of media failure and technological change.

**Digital Provenance**: Information on the origin of a digital object and also on any changes that may have occurred over the course of its life cycle.

**Digitization**: The process of transforming analog material into binary electronic (digital) form, especially for storage and use in a computer. Primarily for access and not preservation.

**Fixity**: A mechanism to verify that a digital object has not been altered in an undocumented manner. Checksums, message digests and digital signatures are examples of tools to run fixity checks. Fixity information, the information created by these fixity checks, provides evidence for the integrity and authenticity of the digital objects and are essential to enabling trust.

**Ingest**: One of the functions listed in the framework for OAIS. It involves taking an object (or objects) into a digital repository.

**Interoperability**: A property of a product or system, whose interfaces are completely understood, to work with other products or systems, present or future, without any restricted access or implementation.
**Migration**: 1. The process of copying data from one type of storage material to another to ensure continued access to the information as the data structure becomes obsolete; media migration. - 2. The process of converting a data from an obsolete structure to a new structure to counter software obsolescence.

**Metadata**: Latin term meaning “information about information.” In the digital realm, metadata is data that describes key information about the digital objects (image files, text files, digital audio/video) and, when appropriate, the original objects they represent. There are different kinds of ‘metadata’ including bibliographic or descriptive metadata, technical metadata, administrative metadata and structural metadata.


**OAIS**: OAIS is an acronym that stands for Open Archival Information System. It is an archival framework developed by the Consultative Committee for Space Data Systems (CCSDS). The OAIS framework consists of an organization of people and systems who have accepted the responsibility to preserve information and make it available for a certain group of people. It does not offer a definitive guideline for how a digital repository should act or what it should do but instead gives the digital preservation community a common language and outlook for talking about digital preservation.

**OAI-PMH**: The Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) is a protocol developed by the Open Archives Initiative. It is used to harvest (or collect) the metadata descriptions of the records in an archive so that services can be built using metadata from many archives.

**Permissions**: The access available to system users attached to specific roles in a computing environment, as well as the mechanism for administering access to a specific object on a computer system. Depending on the system or application, permissions can be defined for a specific user, specific groups of users, or all users; or for a role, or groups of roles; or based on one or more user attributes.

**PREMIS**: Stands for Preservation Metadata: Implementation Strategies. It is now in its second iteration. PREMIS metadata is contained within larger metadata schemas such as METS. PREMIS metadata structures and describes what sort of preservation has been done to a digital object. This might include taking the object into a new archive or changing the format of an object.

**Preservation copy**: Digital content targeted for preservation that is considered the master version of the intellectual content of any arbitrary digital resource. Preservation master files may capture additional information about the original beyond the content itself. Because they are created to high capture standards, preservation master files could take the place of the original
record if the original was destroyed, damaged, or not retained. Preservation masters generally do not undergo significant processing or editing. Preservation masters are often used to make other copies including reproduction and distribution copies.

**Process (noun):** A continuous and regular action or succession of actions occurring or performed in a definite manner, and having a particular result or outcome; a sustained operation or series of operations.

**Process (verb):** To register or interpret (information, data, etc.); Computing to operate on (data) by means of a program

**Refreshment:** Copying a digital object from one media format, such as a CD, to another, such as a hard drive.

**Render:** To make a Digital Object perceptible to a user. This is done through use of a software program and is often used when talking about the emulation of a digital object.

**Schema:** A formal description of a data structure. For XML, a common way of defining the structure, elements, and attributes that are available for use in an XML document that complies to the schema.

**Storage: Archival:** The category of digital storage that provides the services and functions for the long term storage, maintenance and retrieval of digital objects.

**The cloud:** Cloud computing is a type of computing that relies on *sharing computing resources* rather than having local servers or personal devices to handle applications.

In cloud computing, the word cloud (also phrased as "the cloud") is used as a metaphor for "the Internet," so the phrase *cloud computing* means "a type of Internet-based computing," where different services — such as servers, storage and applications — are delivered to an organization's computers and devices through the Internet.

**Trusted Digital Repository:** One whose mission is to provide reliable, long term access to managed digital resources to its designated community, now and in the future. The TDR must include the following seven attributes: compliance with the reference model for an Open Archival Information System (OAIS), administrative responsibility, organizational viability, financial sustainability, technological and procedural suitability, system security, procedural accountability.