Ohio Five Digital Preservation Task Force
Final Report—March 2016

Describe current plans underway among the Five Colleges of Ohio to adopt specific digital preservation systems. What systems are under consideration and how have these been evaluated?

In spring 2016, Denison will begin a one-year pilot of Preservica. Sasha Griffin has identified various types of collections to test, but the highest priority materials include archival digital content and long-term (5+ years) electronic records identified in the university's records management schedule.

Denison will pilot Preservica's Cloud Edition "Starter" subscription plan, which includes up to 250 GB of storage in Amazon S3 and remote training. After the first year, staff will determine whether or not Preservica meets institutional digital preservation needs. If it does, ITS may roll the annual subscription cost into its budget.

In 2014, Oberlin College Library formed a Digital Preservation Subcommittee of the Digital Projects Workgroup. The subcommittee initially considered purchasing a license for DuraCloud, primarily for its high functionality for sound and video files, affordability, and ease of use, as described by Liz Bishoff of the Bishoff Group at a 2014 ARL workshop. The formation of the Ohio Five Digital Preservation Task Force provided the opportunity to investigate more options, so Oberlin has not participated in a DuraCloud pilot.

Currently, Kenyon, Ohio Wesleyan, and Wooster have no plans in process to implement digital preservation systems.

Investigate the advantages and disadvantages of using a single shared digital preservation system for the Ohio Five.

There are currently three prominent vendor-supported digital preservation systems that provide the full range of functionality needed to conform to Trusted Digital Repository (TDR) requirements. These include ArchivesDirect, Preservica, and Rosetta. To learn more about whether or not a shared single instance of a digital preservation system might be a viable option for the Ohio Five, the Task Force spoke with vendors to learn more about the capabilities of these systems.

Conversations with representatives from Artefactual Systems (ArchivesDirect), Preservica, and ExLibris (Rosetta) yielded no examples of a consortium sharing a single instance of a digital preservation system. Vendors expressed the following concerns about adopting that model:

- Performance issues resulting from multiple institutions running processes simultaneously, particularly during the ingest of large files such as video content;
- Confusion caused by sharing a dashboard that displays other institutions' processes;
- Privacy concerns if institutions can see each other's restricted content; and
- Limitations to customizing workflows at the institutional level.

Should the Ohio Five proceed with multiple instances of a common system, the Task Force identified the following advantages:

- Ability to combine training during implementation;
- Opportunity to collaborate on common standards (including metadata); and
- Creation of a community of practice that can share experience.

Note: The Task Force contacted OhioLINK to learn more about the implementation of Rosetta and explore any related opportunities for collaboration. Currently, OhioLINK plans to use Rosetta as a dark archive for EJC and ETD content, and staff is unlikely to discuss the potential for expanded use until after that content has been ingested. OhioLINK has no plans to offer Rosetta licenses to or support instances for member institutions. In the future, OhioLINK may consider allowing individuals from member institutions without digital preservation systems to deposit content.

In addition to ArchivesDirect, Preservica, and Rosetta, the Task Force researched storage-only options APTrust, DuraCloud, and MetaArchive. These systems require the adoption of a separate processing application like Archivematica to prepare content and related metadata for ingest, but all three vendors would allow the Ohio Five to split the costs of an individual institutional membership.

For a more detailed discussion of additional models, please see Potential Approaches to Consortial Digital Preservation on page 5.

Investigate digital preservation initiatives at other liberal arts colleges to understand how they have been implemented and how successful they have been.

The Task Force targeted a list of potential liberal arts colleges to contact based on existing professional relationships, online articles about digital preservation initiatives, director recommendations, and—after initial contact was made—referrals from participating colleges.

In an effort to cover both the technical and staffing issues associated with digital preservation, the Task Force developed the following list of common questions:

1. What system(s) do you use? (may be multiple systems for the following functions: public discovery/access, internal search/file management, digital preservation/file normalization, redundant storage, etc.)
2. How many files are in your system (# of files and size)?
3. What type of content do you preserve?
4. What criteria did you use to evaluate preservation systems?
5. What unforeseen challenges did you encounter during planning and implementation?
6. What kind of technical infrastructure does your institution have in place that supports your digital preservation activities?
7. How much IT support does the library receive? How essential would you say this collaboration was during the implementation phase? During the maintenance phase?
8. What kind of staffing are you able to devote to your digital preservation activities?

Of the six institutions and one consortium interviewed, three (Berea College, Colby-Sawyer College, and Emerson College) implemented Preservica, one (Colgate University) implemented DuraCloud, and three (Amherst College, Grinnell College, and Tri-College Consortium) had yet to select a digital preservation system.

Colleges that implemented Preservica mentioned limited IT support and few staff positions devoted to digital preservation as major reasons they selected a cloud-based, all-in-one solution. Overall, they indicated they were pleased with Preservica's features and responsive technical support.

Colgate uses a RAID array for local storage and access of digital content and selected DuraCloud for long-term storage, scalability, and its support for fixity checking.

Currently, Grinnell has no digital preservation program or plans to adopt a digital preservation system. Amherst uses Fedora 3 but notes that staff efforts are focused on access rather than preservation. In 2015, TriCo's Digital Asset Management and Preservation Working Group (DAMP!) conducted an assessment of digital preservation platforms and determined Hydra/Fedora would work best for the consortium. Because of limited IT support, the working group plans to follow the development of Hydra-in-a-Box and will be participating in a trial of ArchivesDirect.

In almost all cases, librarians and archivists most familiar with the content handled processing and administered the digital preservation systems. IT participated in initiatives by extracting content from older systems, maintaining servers, and vetting service agreements.

Identify any technical and/or institutional barriers we might encounter in implementing a collaborative digital preservation initiative and ways we might mitigate those.

The building of new infrastructure can introduce a number of technical barriers into a collaborative initiative. Locally hosted digital preservation systems come with both upfront and ongoing equipment costs, including hardware purchases (servers and dedicated machines for running digital preservation software) and maintenance. The software itself can be a barrier to a particular collaborative model, as is the case with ArchivesDirect and Preservica's inability to support multiple institutions using single instances of their systems. Adopting cloud-based solutions wherever possible and broadening the collaborative approach to include multiple instances of a common system are ways to address these technical barriers.

Individual needs can present institutional barriers to collaboration. Not all institutions handle restricted content, but those that do have identified specific technological and policy requirements for those materials. One way to mitigate this barrier would be to continue the conversation among the members of the Ohio Five, perhaps in the form of a digital preservation interest group that would discuss issues like these during the system selection and implementation phases.

Identify individuals who are able and willing to contribute time to digital preservation at each institution.

As part of the March 10 retreat with Meg Miner of Illinois Wesleyan University/Digital POWRR, the Task Force developed a list of campus community members who would need to be included in digital preservation initiatives:

- Directors, for their relationships with Provosts and other members of Administration;
- IT Staff, for their technical knowledge and support (security audits, system updates, etc.);
- Educational Technologists and Liaison Librarians, for their existing relationships with faculty;
- Archivists, for their role as curators of records of permanent value;
- Content Creators, including faculty, practitioners/staff, and student workers; and
- Office Staff, as record keepers for their departments.

Using the earlier list as a guide, the Task Force identified the following individuals as potential contributors to digital preservation initiatives at the Ohio Five:

Five Colleges of Ohio
- Ben Daigle, Associate Director of Consortial Library Systems
- TBD, Library Web Services Assistant

Denison University
- Debby Andreadis, Deputy Director
- Sasha Griffin, University Archivist & Special Collections Librarian

Kenyon College
- Abigail Miller, College and Digital Collections Archivist
- Jenna Nolt, Digital Initiatives Librarian

Oberlin College
- Megan Mitchell, Digital Initiatives Coordinator
- Anne Salsich, Associate Archivist
- Jeremy Smith, Special Collections Librarian/Curator of the James and Susan Neumann Jazz Collection

Ohio Wesleyan University
- Emily Gattozzi, Digital Initiatives Librarian

College of Wooster
- Catie Newton, Digital Scholarship & Preservation Librarian

Identify training needs for those who would be participating in digital preservation initiatives.

Training needs depend largely on the requirements of the system(s) implemented. All-in-one systems ArchivesDirect and Preservica include on-site staff training during the implementation stage as part of their subscription plans. In addition to training, ArchivesDirect also offers one-on-one consultation services to help libraries identify the most effective workflows.

Training varies for the three storage-only solutions the Task Force investigated. APTrust expects member institutions to use BagIt to provide packaged content for ingest. Program Director Chip German indicated that APTrust's members generally assist each other with questions about submitting content, although it's possible staff from the organization's central office or a member institution could host a workshop for the Ohio Five.

DuraSpace includes an introductory training session in its subscription fee for DuraCloud and offers as-needed training during the subscription period in instances of staff change, new feature rollouts, etc.

MetaArchive provides skills recommendations for member institutions in its technical specifications document and "regularly" provides member training in those areas.

Implementing an open source system like Archivematica would require the Ohio Five to organize its own training or purchase online training or an on-site workshop from Artefactual Systems. Archivematica's documentation includes manuals for administrators and users, and its online

community features a forum where users can seek help from each other or request support from developers.

In addition to systems-based training, the Task Force expressed an interest in pursuing continuing education and professional development opportunities related to general digital preservation concepts. Members identified Society of American Archivists courses with a technical focus, such as Accessioning and Ingest of Electronic Records, and the International Conference on Digital Preservation as potential options.

Potential Approaches to Consortial Digital Preservation

During the March 10 retreat with Meg Miner, the Task Force determined that not all Ohio Five content warrants full digital preservation in a complete system. Specifically, the group felt the need to distinguish between archival born-digital collections and digitized surrogate content, restricted and open content, permanent and transitory collections, and ongoing and complete collections.

Taking the above into consideration, the Task Force agreed to present the following approaches to digital preservation as potential options for the Ohio Five. (For pricing information, please see the System Comparison Charts on pages 7 and 8.)

1. **All-in-one approach**

   Each institution would have its own instance of an all-in-one system. The Task Force expressed a preference for Preservica, although the group did not rule out ArchivesDirect. Preservica provides a single software solution that would enable institutions to customize processes and set limitations for restricted content.

2. **Hybrid approach:** Digital archival storage (library collections) + full digital preservation system (archives collections)

   Because of the significant differences in institutional needs throughout the Ohio Five, a second approach would address digital preservation at two separate levels: 1) a minimal level for digitized library content (non-originals) and 2) a full level for archival, permanent born-digital content for institutions that provide those archival services. In this approach, the Task Force strongly suggests that archival born-digital content be given top priority. Both of these preservation approaches must be addressed in order to have a workable solution for both libraries and archives.

   The minimal level for digital preservation would include a shared, consortial digital archival storage system that would provide redundant storage of library collections of digitized material, like collections created by the digital scholarship grants. This shared storage system would present a solution for backing up the access collections in an archival storage capacity but would not address a standards-compliant level of digital preservation. The Task Force concluded that this is the only part of the hybrid approach that could be done consortially.

   The full digital preservation level would then be reserved for high priority, archival, born-digital content such as video and audio recordings, born-digital photographs and other electronic records. Each institution with archival programs that accept born-digital content

and/or have electronic records management responsibilities would then be responsible for pursuing its own full digital preservation system for archival and long-term/permanent born-digital materials, since each institution's needs are different.

The advantage of a hybrid approach is that it would allow for some of the cost burden to be shared by the consortial digital archival storage system, as a full archival digital preservation system is not needed to provide services for library collections that only require minimal storage preservation (such as for materials currently in the Ohio Five instance of CONTENTdm). This would allow institutions with archives that are handling high priority and at-risk born-digital content to pursue a full digital preservation system at a lower storage cost. Simultaneously, it would allow institutions that do not need those services to pay only for the level of preservation that is needed for digitized collections accessible through the library.

Options for an all-in-one digital preservation system include Preservica and ArchivesDirect, as mentioned above.

Options for shared digital archival storage include APTrust, DuraCloud, and MetaArchive.

DuraCloud arose as a promising solution for a shared consortial storage solution; it also has the added feature of offering streaming services. All five institutions would share an instance of DuraCloud for content storage. As its name suggests, DuraCloud runs entirely in the cloud. It accepts all formats, although the Ohio Five could collaborate on a consortial standard for pre-ingest processing. On its own, DuraCloud is not OAIS compliant.

These options would require the use of one or more processing applications to prepare the content for ingest and to assign appropriate metadata. For this, Archivematica is a popular option and likely the most fully featured. Archivematica is compatible with the Ubuntu Linux operating system and, if selected, each institution would install Ubuntu on a local computer. This could be installed using VirtualBox, an application that allows users to run multiple operating systems on a single computer. Or, it could be installed as the sole operating system on a dedicated workstation. Prior to proceeding with this approach, the Task Force would need to test performance on a virtual machine against performance on a dedicated workstation.

There are a number of other open source applications that specialize in specific digital preservation tasks. Depending on the minimum standards agreed to by the Task Force, one or more of these applications may also be adopted to prepare content for ingest into the shared digital archival storage system.

3. No consortial approach

Should each institution choose to implement a digital preservation system independently—or not implement a system at all—the Ohio Five would lose an opportunity to build a community of practice where practitioners could collaborate on standards and share expertise. If no system is implemented, institutions risk the loss of many hours of work that have already gone into creating online collections, access to grant funded projects that the Ohio Five agreed to preserve, and irreplaceable content that is born-digital on each campus.

System Comparison Charts

	ALL-IN-ONE APPROACH	
	ArchivesDirect (Archivematica + DuraCloud)	**Preservica (Cloud Edition)**
Storage System	Amazon S3 & Glacier	Amazon S3 & Glacier
Regular Price per Inst	$9,999	$11,900
Consortial Discount Price per Inst	$8,999 (if 3 or more subscribe)	Willing to negotiate if multiple institutions would purchase at the same time, but they have not been able to offer an official quote.
File Storage	1 TB/institution	1 TB/institution
Additional File Storage	1 TB for $825	1 TB of S3 for $1450 1 TB of Glacier for $550
Pros	● All-in-one software + storage option. ● Price includes support, consulting, and training during implementation. ● Based on open source software—capable of developing add-ons/customizations down the road if needed. Also have the option to move from hosted to "on premise" if needed in the future. Flexible. ● Integration with access tools—CONTENTdm, Fedora, DSpace.	● All-in-one software + storage option. ● Price includes support, consulting, and training during implementation. ● Potential for automation of obsolescence and repair of corrupted files. ● Examples of liberal arts institutions with one-person staffs using successfully. ● Year-by-year subscription—option to migrate to another system after first year if needed.
Cons	● Cost? ● Learning curve with the Archivematica application?	● Cost? ● Proprietary system—perhaps not as flexible as an open source option for custom add-ons if needed.
Annual Cost per Inst for 2 TB of File Storage	Regular: $10,824 Discounted: $9,824	Regular (S3): $13,350 Regular (Glacier): $12,450 Discounted: Open to negotiation if multiple institutions purchase at the same time.

	HYBRID APPROACH (STORAGE ONLY)		
	APTrust	**DuraCloud**	**MetaArchive**
Storage System	Amazon S3 & Glacier	Amazon S3 as primary storage. Secondary storage available in Glacier, Rackspace, and San Diego Supercomputer Center	LOCKSS (network of servers distributed across MetaArchive member institutions)
Regular Price per Inst	$20,000	$5,750	$2,500 = membership $5,500 = LOCKSS server
Consortial Discount Price Per Inst	$4,000	$1,150	$1,600
File Storage	10 TB	1 TB	Pay for storage used ($.59/GB/year)
Additional File Storage	5 TB for $2750	1 TB for $500	1 TB for $585
Pros	● Very reliable storage solution with 6 copies distributed across different geo-locations. ● Strong commitment from existing member partners. ● Quarterly fixity checks on all content in storage. ● Strong community—all higher ed institutions trying to solve similar problems and sharing knowledge.	● Runs entirely in the cloud ● Streams audio and video content. ● Able to support a shared single instance for storage. ● Provides dashboard functionality and logs of files that were repaired.	● Up to 7 copies stored across different geographic locations. ● Strong commitment from existing members. ● If consortium could get a large number of participating institutions, costs drop considerably. With 20 institutions, annual cost is $350/year for 2 TB.
Cons	● No good solutions for access yet. ● Only solves the storage problem—not the workflow and processing issue. Still need another tool for that. ● Other members are research institutions operating on a different scale.	● Stores only one copy in S3 by default. Can store additional copies for additional fees.	● Requires that members host staging server. Institutions would need to consider costs associated with server purchase, maintenance, and replacement and investigate any security concerns with campus IT. ● Offers only 3-year membership terms.
Notes	These options are contingent on adoption of a tool to process/transfer/ingest information packages (e.g. Archivematica).		
Annual Cost per Inst for 2 TB of File Storage	$4,000	$2,050	$2,770