**Digital POWRR – Preserving digital Objects With Restricted Resources**  
**IMLS Grant LG-05-11-0156-11**

**A Final Report to the Institute of Museum and Library Services**

**NARRATIVE**  
February 2015  
Jaime Schumacher, Project Director  
jschumacher@niu.edu

**GP:** The notation of GP throughout this report indicates the item appears in the Grant Products section of this packet

**Table of Contents**

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Project Partners</td>
<td>1</td>
</tr>
<tr>
<td>Overview</td>
<td>2</td>
</tr>
<tr>
<td>Changes</td>
<td>2</td>
</tr>
<tr>
<td>Activities Completed During the Project</td>
<td>2</td>
</tr>
<tr>
<td>Activities within the original scope of work</td>
<td>2</td>
</tr>
<tr>
<td>Additional activities performed</td>
<td>7</td>
</tr>
<tr>
<td>Project Results</td>
<td>7</td>
</tr>
<tr>
<td>What’s Next?</td>
<td>10</td>
</tr>
<tr>
<td>Appendix A – Supporting Documentation</td>
<td>12</td>
</tr>
<tr>
<td>A – 1: POWRR Workshop Impact and Outcomes – Additional Supporting Documentation</td>
<td>12</td>
</tr>
<tr>
<td>A – 2: POWRR Tool Grid Impact - Supporting Documentation</td>
<td>16</td>
</tr>
<tr>
<td>A –3: POWRR as a change agent for solution pricing and delivery – Supporting Documentation</td>
<td>18</td>
</tr>
<tr>
<td>A – 4: POWRR as a resource for instructors and practitioners - Supporting Documentation</td>
<td>23</td>
</tr>
<tr>
<td>A – 5: POWRR’s findings as reference point – Supporting Documentation</td>
<td>26</td>
</tr>
<tr>
<td>Appendix B – External Project Summative Evaluation</td>
<td>28</td>
</tr>
<tr>
<td>Appendix C – Conference Presentations, Papers, Posters, Workshops, and Participation</td>
<td>31</td>
</tr>
</tbody>
</table>
Northern Illinois University – Team Members:

<table>
<thead>
<tr>
<th>Name</th>
<th>Position</th>
<th>Name</th>
<th>Position</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lynne M. Thomas</td>
<td>Co-Primary Investigator</td>
<td>Nathan Books</td>
<td>Programmer</td>
</tr>
<tr>
<td>Drew E. VandeCreek</td>
<td>Co-Primary Investigator</td>
<td>TJ Lusher</td>
<td>Administrative Liaison</td>
</tr>
<tr>
<td>Jaime Schumacher</td>
<td>Project Director</td>
<td>Joseph Thomas</td>
<td>Systems Administrator</td>
</tr>
<tr>
<td>Stacey Erdman</td>
<td>Technical Coordinator</td>
<td>Matthew Short</td>
<td>Metadata Librarian</td>
</tr>
<tr>
<td>Danielle Spalenka</td>
<td>Regional History Center Lead</td>
<td>Sarah Fraser</td>
<td>Graduate Assistant</td>
</tr>
<tr>
<td>Katharine White</td>
<td>Regional History Center Lead</td>
<td>Amanda Miller</td>
<td>Graduate Assistant</td>
</tr>
</tbody>
</table>

The team based at NIU contributed to the project by guiding all activities from initiation through completion; administering grant finances, reports, contracts with vendors and consultants, and all deliverables; maintaining the website and wiki; developing the POWRR brand and promoting the project; creating an institution-specific case study (GP); interviewing faculty members to gauge local digital preservation (DP) knowledge and practices; locally installing, testing, and reporting on all DP tools and services; providing the perspective of a mid-sized, state directional university to the Advisory Board as board members guided the direction and activities of the project; attending and presenting at conferences to build awareness of the project’s efforts and outcomes; leading the creation of the POWRR workshop curriculum; scheduling, managing logistical arrangements of, handling registration for, administering evaluative measures on, and teaching the workshop at various venues across the country; authoring the white paper and managing its editorial process; leading the creation of all other grant products, including the Tool Grid.

Chicago State University – Aaisha Haykal and her team from CSU contributed to the project by creating an institution-specific case study (GP); interviewing faculty members to gauge local DP knowledge and practices; locally installing, testing, and reporting on several DP tools and services; providing the perspective of a small, urban, state directional university to the Advisory Board; attending and presenting at conferences to build awareness of the project’s efforts and outcomes; collaborating on the creation of the workshop curriculum; teaching the workshop at various venues across the country; and assisting in the creation of most grant products. The CSU team included Dr. Sharon Hu, Martin Kong, and Gayle Porter.

Illinois State University – Patrice-Andre Prud’homme contributed to the project by creating an institution-specific case study (GP); interviewing faculty members to gauge local DP knowledge and practices; locally installing, testing, and reporting on several DP tools and services; providing the perspective of a mid-sized, state directional university to the Advisory Board; presenting at conferences to build awareness of the project’s efforts and outcomes; collaborating on the creation of the workshop curriculum; teaching the workshop at various venues across the country; and assisting in the creation of most grant products.

Illinois Wesleyan University – Meg Miner contributed to the project by creating an institution-specific case study (GP); interviewing faculty and staff members to gauge local DP knowledge and practices; locally installing, testing, and reporting on several DP tools and services; providing the perspective of a small, private university to the Advisory Board; presenting at conferences to build awareness of the project’s efforts and outcomes; collaborating on the creation of the workshop curriculum; teaching the workshop at various venues across the country; and assisting in the creation of most grant products.

Western Illinois University – Dr. Jeff Hancks contributed to the project by creating an institution-specific case study (GP); interviewing faculty members and local cultural heritage organizations to gauge local digital preservation knowledge and practices; locally testing and reporting on some DP tools and services; providing
Overview

The POWRR project investigated and reported on scalable digital preservation (DP) solutions for small and mid-sized institutions often faced with small staff sizes, restricted IT infrastructures, and tight budgets. Its major deliverable, a white paper, has been well-received and widely read. During the investigation, POWRR uncovered the particular challenges and needs of under-resourced institutions and worked to address and overcome obstacles that often prevent practitioners from taking initial steps in preserving digital content. As a result, POWRR also delivered a well-marked, practical path towards sustainable digital stewardship by: creating a graphic-based tool grid that maps the functionalities of 60+ DP tools/services to an OAIS-based curation lifecycle; developing and teaching a pragmatic, hands-on workshops across the country on the initial steps necessary to accession digital content and how to realistically approach developing a DP program; and producing case studies, collaboration models, and other tools to aid smaller institutions in their DP efforts.

Changes

The following changes were submitted for approval in January, 2013 and approved by Senior Program Officer Charles Thomas on March 5, 2013:

- A 1 year, no-cost extension – Due to the amount of time required to hire the project director (6 months) at the beginning of the award period and the alternative project findings dissemination and education model chosen, additional time was required to successfully complete project goals.
- Funds were reallocated from contractual services and technology supplies (actual expenses for these areas proved lower than original estimates) to project-related travel in order to execute the chosen project findings dissemination and education model.
- The original dissemination and education model called for a national conference to be held on the main campus at Northern Illinois University in DeKalb, IL. After formative evaluation discussions with the Board of Advisors, it was agreed that such a conference in that location would not achieve the critical mass necessary to reach the project’s target audience, especially given the increasingly restricted travel budgets of practitioners at targeted institutions. Instead, the team chose to create a curriculum for a hands-on, full-day workshop that team members then conducted across the country.
- Funds were reallocated to extend the contract of the professional librarian hired as the project director by 6 months. This allowed a key team member to be available during the final 6 months of the project when the majority of dissemination activities took place.

Activities Completed During the Project

Activities within the original scope of work

Below are the high-level activities contained in the original scope of work (presented in **BOLD**), followed directly by the efforts undertaken to complete those activities.

Form an Advisory Board to help shape, oversee and direct the project. Convene initial meeting of Advisory Board, Project Partners, and Project Leadership Team in DeKalb, Illinois to consult and review project goals and budget, and develop guidelines for how options will be critiqued.

Six experienced professionals considered to be experts in the field of digital preservation were selected in
coordination with the IMLS to sit on the POWRR Advisory Board and provide formative guidance and consultation to the project.

Liz Bishoff – Principal Partner, Bishoff Group, LLC
Steve Bromage – Executive Director, Maine Historical Society
Martin Halbert – Dean of Libraries, Associate Professor, University of North Texas
Jerome McDonough – Associate Professor, University of Illinois
Chris Prom – Assistant University Archivist and Professor, University of Illinois
Amy Rudersdorf – Assistant Director for Content, Digital Public Library of America

Draw upon scholars/faculty, from diverse locations and disciplines, to ensure that the project team takes into account faculty needs for creating and using content that is intended to be preserved long term. The POWRR team completed more than 60 campus-wide interviews of faculty, staff, and administrators at NIU and each of the 4 partner institutions to build awareness, gathered digital objects to be used in testing, and collected data on the following:

- The format types and size of digital materials being created
- The methods of storing files and the media instances of data loss and its impact
- Which digital materials are highest priority for recovery in the event of a catastrophic loss
- The confidence levels that data can be recovered if lost and that files will be usable in 25 years
- The desire to have digital material accessible and usable by colleagues in 25 years

The data from the interviews were collected into a database (http://commons.lib.niu.edu/handle/10843/13609) and analysis has revealed interesting and surprising findings (GP).

Conduct an environmental scan of tools and services available to small and medium-sized academic libraries, including cloud-based services like DuraCloud.

The team created a DP tool evaluation rubric based on the OAIS digital curation and preservation lifecycle, and conducted an environmental scan of DP tools, technologies, and services available as options to institutions with fewer resources. The master list of options and the rubric can be found on the wiki at http://powrr-wiki.lib.niu.edu/index.php/Tool_Functionality_Evaluation. For the 60+ tools, services, etc. that were not selected for in-depth testing, the project team used this rubric to create the POWRR Tool Grid, a graphic-based resource for practitioners that shows, at-a-glance, the functionalities of the DP tools and services and how they fit within an OAIS-based digital curation lifecycle (GP).

Select three or four tools and/or services to be subjects of project evaluation. Contact representatives of selected service providers to negotiate terms of service and discuss technical infrastructure required at Northern Illinois University to install effective pilot implementations. Consult with experts and service providers to design pilot implementations of selected tools/services, and then install and maintain pilot implementations at each Partner Institution, creating detailed documentation of work processes. After conducting initial investigations of the tools/services selected by the Advisory Board for inclusion in the study in Year 1, some changes to the list were required. Under advisement of the Board, the team eliminated 2 tools/services from the initial list, added a new one, and altered the approach to the testing of another. Below are the tools/services that were tested by the project team or were eliminated from the testing pool (including the reason).

DuraCloud: Established an account with this hosted service and received the standard training on the use of the service offered to all customers.
MetaArchive: Established a contract with Educopia Institute for a pilot instance of this service. Agreed to pilot a new process, using Bagger, that was believed to be a simpler method of ingesting content into the MetaArchive LOCKSS network. Set up the pilot in a way that mimics a Collaborative Membership with a staging server at Northern Illinois University to which the partner institutions transfer their content for harvesting.

Archivematica: This free, open source tool required no contract. Since part of the evaluative framework for testing these tools/services includes the ease of downloading and installing the software for practitioners within our target audience, the partner institutions attempted to do so without the aid of the project’s technical coordinators. After these attempts were documented, the technical coordinator and project director assisted with the installations as necessary.

Preservica: After receiving an estimate for $28,890 (included 1 year of licensing fees for the software, 1TB of cloud storage, and training) and negotiation attempts with the vendor were rejected, the Advisory Board concluded that this service would not be affordable or scalable to the study’s target audience, particularly considering the quoted rate of $8100 for storing 1TB/year. It was, therefore, removed from the testing list. Towards the end of the project’s testing phase, however, project team members were informed that the company’s pricing and delivery model for this service had changed to more closely reflect the model our team proposed to the vendor in the original failed negotiations. After informing the Board of Advisors of these developments, it was agreed that the team would test Preservica in an abbreviated pilot.

HOPPLA: After initial investigations revealed that HOPPLA had not been updated in some time and parts of the tool did not appear to be functional, the team reached out to the creators of HOPPLA. They stated that they were not maintaining the HOPPLA prototype and that there were no plans to update it. After presenting this development to the Board of Advisors, they voted to remove the tool from the testing list.

Curator’s Workbench: With the removal of both Preservica and HOPPLA from the testing list, we were left with only 1 “processing” type of tool to evaluate. In light of this, the team suggested a handful of potential processing tools to add to the list. The Advisors selected Curator’s Workbench to round out the types of tools/services that the project team would be testing. It is a free, open source tool that requires no contract. Since part of the evaluative framework for testing these tools/services includes the ease of downloading and installing the software for practitioners within our target audience, the partner institutions attempted to do so without the aid of the project’s technical coordinators. After these attempts were documented, the technical coordinator and project director assisted with the installations as necessary.

Internet Archive*: Upon additional investigation by the project team after the Advisory Board recommended its inclusion, it was determined that Internet Archive, while not necessarily a robust DP solution for most mid-sized and smaller institutions, could prove to be effective for the smallest of organizations, like local historical societies. Team members established several pilot projects where very small memory institutions attempted to use Internet Archive as an off-site storage solution for their digital materials.

*Note that the team based these determinations on the traditional Internet Archive service (free) and not its Archive-it subscription service.

Discuss viable methods for bringing smaller and medium-sized institutions into the National Digital Stewardship Alliance.

The project director met with NDSA (National Digital Stewardship Alliance) leadership at the Library of Congress to brainstorm viable methods for bringing smaller and medium-sized institutions into its
membership and potential barriers to entry for those organizations. Of particular note was the possibility of creating regionally-based opportunities to make meeting with other members of the NDSA more feasible for those practitioners from smaller institutions with fewer resources available for traveling. As of 2014, the NDSA has held 2 regionally-based events. As a result of these discussions, 2 POWRR partner institutions, Chicago State University and Northern Illinois University, have joined the NDSA and are active in its working groups.

**Purchase equipment on which project pilot instances are to be installed and tested.**
Due to changes and improvements in the structure of the pilots, the team was able to use existing equipment and it was not necessary to purchase additional equipment.

**Keep the public and other institutions apprised of our progress on this project through a project blog, website, and possibly other social media such as Facebook or Twitter.**
A project website (http://digitalpowrr.niu.edu/), wiki (http://powrr-wiki.lib.niu.edu/index.php/Main_Page), and blog (http://digitalpowrr.niu.edu/blog/) were created and updated throughout the project (GP). Project activities were publicized on listservs and Google groups, and team members promoted the project on Facebook and Twitter.

**Develop an evaluative framework for testing and evaluating tools and services, and formally test and evaluate them according to developed guidelines of pilot instances for selected tools and services.**
The project team and the Board of Advisors jointly developed the evaluative framework for the testing and evaluation of the final selection of DP tools and services. It was agreed that the testing should focus largely on the accessibility and usability of the tools/services by a practitioner at an under-resourced institution and the constraints they could possibly be working within such as:

- Outdated technical infrastructure; Limited personal technical skills
- Little to no budget for licensing fees, additional equipment, etc.
- No programmers on staff and no access to server administrators
- No data, metadata, or digital collections librarians on staff

Given the wide range of technical, personnel, and administrative structures at each of the 5 partner institutions, it was agreed to test most of the tools/services at all locations. Project leads also recruited additional people to participate at their institutions to apply a variety of skill sets to the testing process, including archivists, manuscript processors, graduate assistants, etc. Several operating systems were used in the testing phase, but all institutions tested the same version of the tools/service, where applicable. The evaluative framework and raw data from testing can be found on the project wiki at http://powrr-wiki.lib.niu.edu/index.php/Tool_Functionality_Evaluation and the results are, of course, summarized in the project’s primary deliverable, the white paper.

**Discuss potential business models for future implementation of digital preservation solutions for small and mid-sized institutions, both within the Advisory Board and in concert with select service providers.**
The discussion of potential business models for the project’s target audience centered on the following belief: while small and mid-sized institutions do indeed have constraints and fewer resources, they also have unique advantages that can be exploited to full advantage when selecting a business model. For example, an institution with a small number of professional staff can often be more agile and can accelerate decision-making processes. To that end, the team presented four potential solution models in its white paper.

The team agreed that one attractive business model for the project’s target audience is the use of a hosted solution that fulfills the functionality of the greater majority of the digital curation lifecycle. At the time, there
was only one such solution within reach of the target audience. However, its fees were not openly available and the quote POWRR received included storage costs that were seemingly inflated. Therefore, the team suggested an alternative pricing and delivery model (See Appendix A-3).

In testing MetaArchive within a Collaborative Membership model, the team recognized that the development of appropriate organizational structures and legal relationships between collaborating institutions can be a significant barrier to smaller organizations seeking higher levels of DP, due in part to the cost of legal counsel required for the drafting of original agreements. The team commissioned a lawyer, who is also a librarian, to create the business model and legal framework for this type of partnership. The deliverable created in this endeavor will be publicly available under a Creative Commons License and provided in a modifiable format for practitioners to customize as needed. (GP)

Make final decisions as to content of preliminary final report. Determine which solutions are viable, plausible, and cost-effective. Issue formal report on findings and make recommendations on suitable approaches (and accompanying considerations of each approach).

In concert with the Board of Advisors and all Project Partners, the team produced its report in the form of a white paper (GP) entitled “From Theory to Action: “Good Enough” Digital Preservation Solutions for Under-Resourced Cultural Heritage Institutions” that can be found at http://commons.lib.niu.edu/handle/10843/13610.

Hold national IMLS-cosponsored conference on NIU campus, at which project partners and participants present and discuss research and findings to an audience of representatives of small and medium-sized institutions searching for digital preservation solutions.

After formative evaluation discussions with the Board of Advisors, it was agreed that such a conference in DeKalb, IL would not achieve the critical mass necessary to reach the project’s target audience, especially given the increasingly restricted travel budgets of practitioners at targeted institutions. Instead, the team chose to create a curriculum (GP) for a hands-on, pragmatic full-day workshop that team members then conducted across the country.

Outside evaluator reviews and assesses project work.

At the suggestion of the grant’s program officer, the team used a “committee of visitors” approach to this evaluation and assessment. The team invited Amy Rudersdorf, Assistant Director for Content at the Digital Public Library of America, to NIU to meet with team members, access all project documentation and products, and interview Project Partners. Ms. Rudersdorf created a report that was then sent to Dr. Martin Halbert for review. The report can be found in Appendix B.

Primary Investigators will present their findings at appropriate national conferences, and submit discussions of findings for publication in the appropriate professional venues.

The POWRR project quickly became renowned for its efforts and its approach to digital preservation. Its team members were in demand to speak at conferences, conduct workshops, and present papers/posters (GP). Therefore, the majority of the POWRR team, and not just the Primary Investigators, actively participated in the dissemination of project findings. A full list of these activities can be found in Appendix C.

In the 2 years following the end of the grant, project members will convene 4 webinars and offer email support for representatives of institutions considering implementing the project’s resource model.

These activities are ongoing.
Additional activities performed
In the course of the project, the team recognized that practitioners at smaller organizations require specialized and discrete assistance with communicating the need for and benefits of a DP program to key stakeholders within their organization. POWRR chose to partner with a communications consultant to create 1-page handouts for practitioners to use in these efforts that are targeted to specific stakeholders like Academic Administrators, Content Creators, IT Administrators, etc. These documents are publicly available under a Creative Commons License and provided in a modifiable format for practitioners to customize as needed (GP).

The team worked with very small memory organizations like historical societies to test Internet Archive as a DP option. These organizations are typically run almost entirely by volunteers, many of whom are not comfortable working with cloud-based services. Based on the experience and feedback the team received, a user-friendly, screenshot-based tutorial was created to walk a user at this type of institution through the entire process of using Internet Archive for the purpose of digital preservation (GP).

In the course of the investigation, the POWRR team discovered that most of the preservation tools and services available can only be used after initial accessioning and inventorying activities are undertaken, referred to by the team as “triage” activities. The team discovered a viable, open-source tool for this purpose, Duke Data Accessioner, though it had not been updated in some time. POWRR commissioned its original creator to revise the tool making it institution and platform agnostic, creating the capability to add Dublin Core metadata, and creating functionality for transforming the resulting xml report into a format that is human-readable and able to be analyzed (GP).

Project Results
This multi-institutional project focused on investigating scalable DP solutions for small and mid-sized institutions that are often faced with small staff sizes, restricted IT infrastructures, and tight budgets. These institutions hold unique digital content important to their region's cultural heritage, yet many of the practitioners are unsure how to approach the stewardship of the content and are overwhelmed by the large number of DP tools/services available. As the project progressed, the team uncovered the particular challenges, advantages, needs, and desires of under-resourced institutions. They worked to address and overcome obstacles that often prevent practitioners from taking even initial steps in preserving their digital content. POWRR sought to create a well-marked, realistic path towards sustainable digital stewardship for this often overlooked group and, to do so, went beyond their original directive to investigate and report on a handful of DP tools/services.

POWRR created a pragmatic, hands-on, full day workshop to teach the initial steps necessary to accession and inventory digital content as well as how to realistically approach developing a DP program. Recognizing that many of their target institutions currently have little-to-no travel and training budgets, the POWRR team traveled across the country to conduct these workshops for very little cost to the practitioners. 165 practitioners from 138 organizations located in 27 states and the District of Columbia developed crucial digital preservation skills at these workshops.

To gauge each attendee’s knowledge about digital preservation and his/her confidence in his/her abilities to perform digital preservation activities, the team administered tests immediately before and after each workshop. The results in the graph below indicate a significant increase in both knowledge and confidence.
To measure the impact the POWRR workshops had on these practitioners and their institutions, the team surveyed attendees 3 months after their particular workshop. 139 attendees received surveys and 71 responded, a 51.08% response rate. Of those responding, 69.01% indicated that they had indeed initiated digital preservation activities at their institutions in the 3 months following the workshop. Of these, 74.42% of attendees stated that the skills and knowledge gained during the POWRR workshop were very helpful or crucial in choosing and implementing the digital preservation activities. The graph below details which activities the practitioners had initiated at their institutions.

**GRAPH DETAILS:** Each workshop participant completed a workshop pre-test and post-test to measure the skills and knowledge conveyed throughout the workshop. Questions focused on the participant’s level of understanding, level of awareness, ability to build awareness, and ability to perform initial activities. For all questions, answer statements were provided along with a numeric value for participants to select the most relevant statement to his/her situation. Numeric values ranged from 1, indicating low understanding or ability, to 5, indicating high understanding or ability.

**Question 1:** In addressing the participant’s level of understanding, the provided responses ranged from 1: I am unfamiliar with DP tools/services and unsure how they might fit my institutional needs, to 5: I can explain to a colleague how certain tools/services can satisfy different needs within my institution.

**Question 2:** In addressing the participant’s level of awareness and skills with DP tools/services, the provided responses ranged from 1: I am unfamiliar with DP tools and how they work, to 5: I am able to use at least one DP tool with confidence.

**Question 3:** In addressing the participant’s ability to build awareness around DP, the provided responses ranged from 1: I don’t know how to build awareness regarding DP at my institution and am unaware of resources available to help me with this, to 5: I have the resources and enough skills necessary to build awareness at my institution, and have a plan of action for initializing/continuing discussions with others at my institution.

**Question 4:** In addressing the participant’s ability to perform initial pre-ingest DP activities, the provided responses ranged from 1: I don’t understand what pre-ingest activities are, to 5: I have a robust pre-ingest workflow that I can successfully use.
Please see Appendix A-1 for additional supporting documentation relating to the outcomes and impact of the POWRR Workshops.

Knowing that these institutions can achieve economies of scale by working together, POWRR produced collaboration models and the underlying legal framework often needed for these endeavors, all directed at small and mid-sized institutions. These models and framework are currently being used as the basis for forming a Digital Preservation Cooperative in the state of Illinois and are openly available for use by others.

POWRR delivered a well-received, graphic-based tool grid that shows, at-a-glance, the functionalities of over 60 DP tools and services and how they fit within an OAIS-based digital curation lifecycle. The POWRR Tool Grid was the subject of multiple blogs, tweets, etc. including from the Library of Congress’s National Digital Information Infrastructure and Preservation Program, The American Library Association, and Lyrasis. Please see Appendix A-2 for supporting documentation. As use of the POWRR Tool Grid increased and became more widespread, the team was contacted by representatives of an organization in the UK working to consolidate the many DP tool registries available on the Web into a single registry, COPTR, a Community Owned digital Preservation Tool Registry. Knowing that the POWRR Tool Grid was merely a snapshot in time and its upkeep not sustainable by the grant, the team chose, along with several notable institutions, to support this effort by providing them with an export of the Tool Grid data. Given the popularity of the graphic-based look and feel of the Tool Grid, the POWRR team commissioned the steward of COPTR to combine the form and function of the original POWRR Tool Grid with the far greater coverage of tools and sustainability provided by the COPTR data feed, resulting in the POWRR Tool Grid v2 (GP). [http://www.digipres.org/tools/about/](http://www.digipres.org/tools/about/)
POWRR helped to define a market segment of smaller institutions interested in acquiring vendor-based DP solutions. With smaller budgets and the need for scaled-down solutions to reflect smaller digital holdings, this market niche was somewhat overlooked based on the pricing and marketing strategies existing in the early stages of the project in 2012. POWRR advocated on behalf of those institutions and suggested to select DP-solution vendors that they provide scaled-down and transparent pricing geared towards smaller institutions, with notable results. Solution models that accommodate the particular nuances and challenges of this market segment are now available from a variety of vendors. See Appendix A-3 for supporting documentation.

The Digital POWRR Project became a go-to source for practitioners attempting to implement digital preservation programs, as evidenced by the chatter surrounding the project on social media and blogs. POWRR also became a resource for digital preservation instructors and learners, as evidences by its inclusion in the curriculum of esteemed courses like The Society of American Archivists’ Digital Archives Specialist (DAS) Curriculum and Certificate Program. Please see Appendix A-4 for supporting documentation.

POWRR’s findings and its digital preservation functionality grid (below) became a crucial reference point by which vendors and solution-seekers characterize and understand the capabilities of various digital preservation tool and services. The POWRR white paper (GP) alone has been accessed more than 5700 times in its first 6 months of release. Please see Appendix A-5 for supporting documentation.

![Functionality Grid](image)

**What’s Next?**

One of the major challenges of this project was evaluating how the various DP tools and services handled metadata. The tools POWRR investigated often have an automated process for ingesting metadata that has already been created, and the ability to automatically extract or create metadata manually, to different degrees. The team approached testing with “from scratch” collections to explore the functionality of the extant metadata creation options in the tools as they existed, as it was thought best to begin with what the tools did as standalones before seeing how they interacted with other tools. Although it was out of scope for this particular project, the team was nonetheless eager to explore how these tools interfaced with descriptive tools (such as ARCHON or ContentDM) that the partner organizations might already be using. Unfortunately, technical challenges with installations and end-user testing of the software took up so much time that the team was unable to pursue this. Many of POWRR’s followers have inquired about tool and service interoperability and the team believes that a concerted investigation would be a worthwhile endeavor.
Many of the Partners of POWRR are eager to capitalize on the findings of their investigation and move the community of practice they established towards the implementation of a viable DP solution. As awareness of POWRR’s efforts grew, the PI’s were contacted by practitioners from the largest university in Illinois who were interested in collaborating on a potential solution. This solution would take advantage of POWRR’s established community, its outreach capabilities, and its proven educational model, and combine them with the computing power and technical expertise of the larger university into a cooperative that would be open to organizations of all sizes and persuasions. This ongoing endeavor is the basis of a forthcoming application to a federal granting agency.

POWRR conducted its first workshop in the spring of 2014 and its value to practitioners and their institutions quickly become apparent. Several factors convinced the team that this particular deliverable of the project should be sustained after the end of the granting period. Those factors included:

- The overwhelmingly positive direct feedback of participants
- The quantitative evidence of significant impact at institutions
- The frequent inquiries regarding potential workshop dates
- Many workshops would fill to capacity and the wait lists were lengthy

The team applied for and was awarded a major grant from the National Endowment for the Humanities: Celebrating 50 years of Excellence. The project, *From Theory to Action: Extending the Reach of Digital POWRR Preservation Workshops* will allow the team to update, develop, and present workshops to archivists, librarians, and other cultural heritage professionals, aimed particularly at those from small and medium-sized institutions. This award runs from January 2015 through December 2016.
A – 1: POWRR Workshop Impact and Outcomes – Additional Supporting Documentation

Selected comments from POWRR Workshop Evaluations:

› SO HAPPY to be moving beyond theory and into the more practical realm.
› One of the most helpful workshops I've been to in a number of years.
› All the instructors are approachable, clear, well-organized, helpful, articulate, practical, experienced and willing to share their knowledge. This is the first 100% helpful, realistic workshop on DP I've ever taken - and I've taken a LOT.
› Thank you! I have a much better understanding of the overall picture of digital preservation activities, workflows, and tools/searches, and how they all fit together. Will be VERY helpful as Archivist of DC/SLA.
› Thank you for shifting focus from strictly academic institutions to include the others.
› I am new to the language and process of digital preservation and now feel completely comfortable coming to others to advocate this.
› The initial step into digital archiving can be scary and seem like such a large task. This seminar showed some very useful tools that can be used to make a start. Seems within reach now.
› Thanks for the informative workshop! Since this particular topic wasn't covered in my archive school program this workshop was a great introduction to digital preservation.
› Thank you! As a small shop I often feel left out of big discussions and your focus on all of us is really meaningful and genuinely helpful to me.
› The handouts and resources listed within were the most valuable part of the workshop. I didn't realize the level of detail involved in such a project. Definitely have a new appreciation.

Workshop participants represented 165 institutions from 27 states and the District of Columbia:
<table>
<thead>
<tr>
<th>Alabama</th>
<th>Illinois</th>
</tr>
</thead>
<tbody>
<tr>
<td>Archives and History – Montgomery, AL</td>
<td>Fr. Michael L. Pfleger Archives – Chicago, IL</td>
</tr>
<tr>
<td>California</td>
<td>Donate Library – Chicago, IL</td>
</tr>
<tr>
<td>California Institute of Technology – Pasadena, CA</td>
<td>Loyola University Chicago – Chicago, IL</td>
</tr>
<tr>
<td>City of College of San Francisco – San Francisco, CA</td>
<td>Molex – Lisle, IL</td>
</tr>
<tr>
<td>Golden Gate National Recreation Area – San Francisco, CA</td>
<td>North Park University – Chicago, IL</td>
</tr>
<tr>
<td>Phoebe Hearst Museum of Anthropology – Berkeley, CA</td>
<td>San Francisco Maritime Nat’l Historical Park – Oakland, CA</td>
</tr>
<tr>
<td>San Francisco Public Library – San Francisco, CA</td>
<td>Sigma Alpha Epsilon – Evanston, IL</td>
</tr>
<tr>
<td>Santa Rosa Junior College – Petaluma, CA</td>
<td>University of Chicago Library – Chicago, IL</td>
</tr>
<tr>
<td>University of California Berkeley – Berkeley, CA</td>
<td>University of San Francisco – Sacramento, CA</td>
</tr>
<tr>
<td>University of San Francisco – Oakland, CA</td>
<td>Alpha Kappa Psi – Indianapolis, IN</td>
</tr>
<tr>
<td>California Digital Library – Oakland, CA</td>
<td>Indiana University South Bend – South Bend, IN</td>
</tr>
<tr>
<td>CalTech – New York, NY, CT</td>
<td>Alpha Gamma Delta – Indianapolis, IN</td>
</tr>
<tr>
<td>Notre Dame University Psychology Dept – Belmont, CA</td>
<td>Kappa Alpha Theta – Indianapolis, IN</td>
</tr>
<tr>
<td>San Francisco Art Commission – San Francisco, CA</td>
<td>National Model Aviation Museum – Muncie, IN</td>
</tr>
<tr>
<td>San Francisco Public Library – San Francisco, CA</td>
<td>Purdue University – West Lafayette, IN</td>
</tr>
<tr>
<td>University and Jepson Herbaria – Oakland, CA</td>
<td>Zeta Tau Alpha – Indianapolis, IN</td>
</tr>
<tr>
<td>University of California Davis – Davis, CA</td>
<td>Pi Beta Phi – Indianapolis, IN</td>
</tr>
<tr>
<td>Yerba Buena Center for the Arts – San Francisco, CA</td>
<td>Theta Chi Fraternity – Carmel, IN</td>
</tr>
<tr>
<td>Connecticut</td>
<td>Kansas</td>
</tr>
<tr>
<td>Iota Imaging – Woodbridge, CT</td>
<td>Fort Hays State University – Canton, KS</td>
</tr>
<tr>
<td>Quinnipiac University – Hamden, CT</td>
<td>Berea College – Berea, KY</td>
</tr>
<tr>
<td>University of St Joseph – West Hartford, CT</td>
<td>Phi Gamma Delta – Lexington, KY</td>
</tr>
<tr>
<td>Wesleyan University – Middletown, CT</td>
<td>Maryville College – Elkhart, IN</td>
</tr>
<tr>
<td>Enfield Public Library – Enfield, CT</td>
<td>Phi Sigma Sigma – Elkhart, IN</td>
</tr>
<tr>
<td>Olin Memorial Library – Middletown, CT</td>
<td>Massachusetts Historical Society – Boston, MA</td>
</tr>
<tr>
<td>U.S. Coast Guard Academy – New London, CT</td>
<td>Brandeis University – Waltham, MA</td>
</tr>
<tr>
<td>Welles-Turner Memorial Library – Glastonbury, CT</td>
<td>Framingham State University – Framingham, MA</td>
</tr>
<tr>
<td>Georgia</td>
<td>Maryland</td>
</tr>
<tr>
<td>Alpha Delta Pi – Atlanta, GA</td>
<td>Massachusetts Historical Society – Boston, MA</td>
</tr>
<tr>
<td>Hawaii</td>
<td>New England Historic Genealogical Society – Boston, MA</td>
</tr>
<tr>
<td>Hickam Field, Joint Base Pearl Harbor – Hickam AFB, HI</td>
<td>Phillips Academy – Andover, MA</td>
</tr>
<tr>
<td>Illinois</td>
<td>Illinois</td>
</tr>
<tr>
<td>Archives of the NA Province of the Cenacle – Chicago, IL</td>
<td>State Library of Massachusetts – Boston, MA</td>
</tr>
<tr>
<td>Black Metropolis Research Consortium – Chicago, IL</td>
<td>Trustees of Reservations Bartholomew’s Cobble – Sheffield, MA</td>
</tr>
<tr>
<td>Chicago State University – Chicago, IL</td>
<td>Western New England University – Springfield, MA</td>
</tr>
<tr>
<td>DePaul University – Chicago, IL</td>
<td>American Jewish Historical Soc. NE Archives – Boston, MA</td>
</tr>
<tr>
<td>DuSable Museum of African American History – Chicago, IL</td>
<td>Boston Architectural College – Boston, MA</td>
</tr>
<tr>
<td>Benedictine University – Lisle, IL</td>
<td>First Parish in Brookline – Brookline, MA</td>
</tr>
<tr>
<td>Catholic Theological Union – Chicago, IL</td>
<td>Isabella Stewart Gardner Museum – Boston, MA</td>
</tr>
<tr>
<td>Chicago Theological Seminary – Chicago, IL</td>
<td>Mount Holyoke College – South Hadley, MA</td>
</tr>
<tr>
<td>Dominican University, Priory Campus – River Forest, IL</td>
<td>NFPA – Quincy, MA</td>
</tr>
<tr>
<td>ELCA Archives – Elk Grove Village, IL</td>
<td>Smith College – Northampton, MA</td>
</tr>
<tr>
<td>Illinois State University – Normal, IL</td>
<td>Suffolk University – Boston, MA</td>
</tr>
<tr>
<td>Meadville Lombard Theological School – Chicago, IL</td>
<td>University of Massachusetts – Northampton, MA</td>
</tr>
<tr>
<td>Morrison Shearer Foundation – Northbrook, IL</td>
<td>Michigan State University – East Lansing, MI</td>
</tr>
<tr>
<td>Northwestern University Library – Evanston, IL</td>
<td>Michigan State University – East Lansing, MI</td>
</tr>
<tr>
<td>Robert M. Myers Archives – Techny, IL</td>
<td>Michigan State University – East Lansing, MI</td>
</tr>
<tr>
<td>The United Library – Evanston, IL</td>
<td>Michigan State University – East Lansing, MI</td>
</tr>
<tr>
<td>State</td>
<td>Institution</td>
</tr>
<tr>
<td>-------------</td>
<td>-----------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Minnesota</td>
<td>Cannon Falls Military History Museum – Cannon Falls, MN</td>
</tr>
<tr>
<td></td>
<td>Historical Society – Owatonna, MN</td>
</tr>
<tr>
<td></td>
<td>Minnesota Discovery Center – Chisholm, MN</td>
</tr>
<tr>
<td></td>
<td>Pope County Museum &amp; Historical Society – Glenwood, MN</td>
</tr>
<tr>
<td></td>
<td>Rice County Historical Society – Faribault, MN</td>
</tr>
<tr>
<td></td>
<td>Stevens County Historical Society – Grand Rapids, MN</td>
</tr>
<tr>
<td></td>
<td>Canby Depot Museum – Canby, MN</td>
</tr>
<tr>
<td></td>
<td>Heritage Preservation Board of Edina – Edina, MN</td>
</tr>
<tr>
<td></td>
<td>Lyle Historical Society – Lyel, MN</td>
</tr>
<tr>
<td></td>
<td>Northfield Historical Society – Northfield, MN</td>
</tr>
<tr>
<td></td>
<td>Renville County Historical Society Museum – Morton, MN</td>
</tr>
<tr>
<td></td>
<td>Sisters of Saint Joseph – Saint Paul, MN</td>
</tr>
<tr>
<td>Missouri</td>
<td>American Truck Historical Society – Kansas City, MO</td>
</tr>
<tr>
<td></td>
<td>Linda Hall Library – Kansas City, MO</td>
</tr>
<tr>
<td></td>
<td>Unity Library and Archives – Lees Summit, MO</td>
</tr>
<tr>
<td></td>
<td>Kansas City Art Institute – Kansas City, MO</td>
</tr>
<tr>
<td></td>
<td>Mathematics Dept Southwest Baptist Uni – Bolivar, MO</td>
</tr>
<tr>
<td>Nebraksa</td>
<td>University of Nebraska at Omaha – Omaha, NE</td>
</tr>
<tr>
<td>New Carolina</td>
<td>Dartmouth College – Hanover, NH</td>
</tr>
<tr>
<td></td>
<td>Phillips Exeter Academy – Exeter, NH</td>
</tr>
<tr>
<td>New York</td>
<td>National Multicultural Greek Council – New York, NY</td>
</tr>
<tr>
<td></td>
<td>The Museum at the Fashion Institute of Technology – New York, NY</td>
</tr>
<tr>
<td></td>
<td>Sigma Sigma Rho – Queens, NY</td>
</tr>
<tr>
<td>North Carolina</td>
<td>North Carolina University Caplain – Raleigh, NC</td>
</tr>
<tr>
<td></td>
<td>Davidson College – Davidson, NC</td>
</tr>
<tr>
<td>Ohio</td>
<td>Kappa Kappa Gamma – Columbus, OH</td>
</tr>
<tr>
<td></td>
<td>Ohio University – Athens, OH</td>
</tr>
<tr>
<td></td>
<td>Oregon</td>
</tr>
<tr>
<td></td>
<td>Oakland Museum – Oakland, OR</td>
</tr>
<tr>
<td>Rhode Island</td>
<td>John Hay Library – Providence, RI</td>
</tr>
<tr>
<td></td>
<td>Redwood Library and Athenaeum – Newport, RI</td>
</tr>
<tr>
<td></td>
<td>Rhode Island School of Design – Providence, RI</td>
</tr>
<tr>
<td></td>
<td>Providence Public Library – Providence, RI</td>
</tr>
<tr>
<td></td>
<td>Rhode Island College – Providence, RI</td>
</tr>
<tr>
<td></td>
<td>Rhode Island State Archives – Providence, RI</td>
</tr>
<tr>
<td>South Carolina</td>
<td>South Carolina State Library – Columbia, SC</td>
</tr>
<tr>
<td></td>
<td>Clemson University – Clemson, SC</td>
</tr>
<tr>
<td></td>
<td>South Carolina State Library – Columbia, SC</td>
</tr>
<tr>
<td></td>
<td>College of Charleston – Charleston, SC</td>
</tr>
<tr>
<td></td>
<td>University of South Carolina – Columbia, SC</td>
</tr>
<tr>
<td>Tennessee</td>
<td>Temple Israel – Memphis, TN</td>
</tr>
<tr>
<td>Nebraska</td>
<td>Kappa Delta Sorority – Memphis, TN</td>
</tr>
<tr>
<td></td>
<td>Texas</td>
</tr>
<tr>
<td></td>
<td>Delta Delta Delta Sorority – Arlington, TX</td>
</tr>
<tr>
<td>New Hampshire</td>
<td>University of Vermont – Burlington, VT</td>
</tr>
<tr>
<td>Vermont</td>
<td>University of Vermont – Burlington, VT</td>
</tr>
<tr>
<td></td>
<td>Virginia</td>
</tr>
<tr>
<td></td>
<td>Sigma Nu Fraternity – Lexington, VA</td>
</tr>
<tr>
<td>Wisconsin</td>
<td>Chippewa Valley Museum – Eau Claire, WI</td>
</tr>
<tr>
<td></td>
<td>Lawrence University – Appleton, WI</td>
</tr>
<tr>
<td></td>
<td>University of Wisconsin – Madison, WI</td>
</tr>
<tr>
<td>Washington DC</td>
<td>American College of Surgeons – Washington D. C.</td>
</tr>
<tr>
<td></td>
<td>Georgetown Law Library – Washington D.C.</td>
</tr>
</tbody>
</table>
3 month follow up survey questions:

Q2 Which workshop did you attend?

- Pilot: Chicago State University: April 2014
- MAC: Kansas City, MO: April 2014
- Data Driven: Charleston, SC: June 2014
- AASLH: St. Paul, MN: September 2014
- San Francisco, CA: September 2014
- Smith College: October 2014
- Northwestern University: November 2014

Q3 In the three months following the workshop, did you initiate any digital preservation activities at your institution?

- Yes
- No

IF YES, THEN THE FOLLOWING QUESTIONS WERE ASKED:

Q4 Which activities did you initiate?

- Began building awareness of DP issues through informal or formal discussions with colleagues and/or administrators
- Initiated/Continued inventorying digital materials
- Enhanced metadata of current records
- Reviewed current collection development policies to assess their applicability to digital materials
- Investigated tools and services presented in the POWRR workshop
- Reviewed other institutions' DP policies
- Engaged in outreach/education activities
- Used Data Accessioner or other accessioning tool
- Read the POWRR White Paper
- Visited the 'I Survived' section of the POWRR website to look for additional resources
- Other ____________________

Q4a How helpful were the skills and knowledge you gained at the POWRR workshop in choosing and implementing the above activities?

- Not helpful
- Helped a little
- Helpful
- Very helpful
- Crucial

IF NO, THEN THE FOLLOWING QUESTIONS WERE ASKED:

Q4 What prevented you from initiating your intended activities?

Q5 What could the workshop have done differently to better assist you?

Q6 Thank you for your feedback! Do you have any other comments?
A – 2: POWRR Tool Grid Impact - Supporting Documentation

Number of POWRR Tool Grid page views from its rollout in April 2012 through Jan 2015: 10,980.

https://twitter.com/ndiipp/status/289049854511046656

https://twitter.com/tjowens/status/448795859061604352

https://www.facebook.com/LyrDigPres/posts/515830055131582
2. Institutions looking to regularly monitor material outside a digital preservation program -- if you have a file store outside of a repository system that needs regular monitoring but you still want to regularly monitor checksums, these might be good options:
   a. Pasty (Windows only for the time being)
   b. Auditing Control Environment
   c. Manually generate/verify checksums (works best for small amounts of data)
   d. HDF5
   e. VerSys File Integrity Monitoring System
   f. Fakty_checker
   g. Fakty
   h. Make your own custom tool (Python)

3. Resources for finding tools (and more information about specific tools):
   a. POWRR project
   b. COPTF wiki (specifically, the fakty category)
   c. CAROL's evaluation of checksum programs (from my very own institution, offers overviews of several free checkers that do not have scheduling capabilities)

And More…

https://www.linkedin.com/groups/This-looked-interesting-useful-Digital-3026810.S.5939003967622119425
http://resources.ethnosproject.org/preserving-digital-objects-restricted-resources-powrr/
http://acrl.ala.org/dh/2013/05/15/resource-digital-preservation-tool-grid/
A –3: POWRR as a change agent for solution pricing and delivery – Supporting Documentation

The following is a memo from the POWRR Project Director to the project’s contact at Tessella outlining a plan to make products like Preservica more affordable and scalable for small and mid-sized institutions. In telephone conferences, the team also lobbied for transparent pricing, so practitioners do not have to solicit quotes from vendors, an often time-consuming process.

Jaime Schumacher

From: [redacted]@tessella.com
Sent: Monday, November 19, 2012 6:39 PM
To: [redacted]
Cc: [redacted]
Subject: Re: Follow up from conference call

Thanks, Jamie, This looks like a great write up on the topics and I'm working on a time to discuss with my colleagues.

I hope you have a great thanksgiving break.

best regards,

Mike

“Jaime Schumacher” <jSchumacher@niu.edu>
11/15/2012 09:09 AM
Subject: Follow up from conference call

Hi Mike,

Thank you again for taking the time to speak with our team. Prior to our call with you, our team had spent some time brainstorming ways that the Preservica product could become more affordable to institutions similar to those carrying out this study. We think that we may have come up with some good ideas and we touched on them in our call with you. Those ideas are outlined later in this message.

What we are essentially looking at here is this: If Tessella thinks that there is the potential to implement some of these ideas to create an alternate pricing model, then Preservica may be affordable to a portion of our target audience as an end-to-end, turnkey, digital preservation solution. If that is the case, then we would be interested in conducting a 12 month pilot of Preservica for the purposes of our study. However, to do that pilot, we would need to work with you on the cost of the pilot (details below).

Or, to put it another way, we are looking at 2 different negotiations here: 1. A negotiation around the current pricing model of Preservica that is offered to institutions of our nature in general. 2. A negotiation to make a pilot instance of Preservica affordable within our grant’s budget. IF we can get to a place where both Tessella and our team agrees that there CAN be some sort of redefinition of the pricing model offered to institutions in general, based possibly on some of the ideas below, THEN we can proceed to the pilot instance negotiation. We understand that your original pricing model would be the foundation for negotiating a pilot instance, as there is no feasible way to implement any of our ideas in the short time frame of our study.

With that said...here is a recap of our ideas:
GENERAL PRICING MODEL IDEAS

License Fees: Currently $11,800 for one year.
How about offering a consortia pricing model? Then the fee could be spread out over several smaller institutions with one institution being the point of contact for Tessella?
How about offering a significantly lower licensing fee to smaller institutions generating less traffic on the hosted service?

Setup and Initial Training Fees: Currently $7,000.
How about allowing multiple institutions in the same geographic area attend the training at one site, thereby sharing this cost?
How about offering remote, real-time training at a reduced hourly fee (by eliminating the trainer’s traveling costs)?
How about creating static, computer-based training modules that can be purchased for a flat fee?

Subsequent Telephonic Training Sessions: Currently $2000 for 4 sessions.
How about allowing multiple institutions to conference in to telephonic training sessions?
How about creating static, computer-based follow-up training modules that can be purchased for a flat fee?
How about creating message boards/on-line help services where users can share knowledge, tips, etc. with each other or pose questions to Tessella experts in open forums.

Storage Costs: Currently $8,100 for 1TB for 1 year with cloud storage provided in 1 of Amazon’s geographical “zones”. **We understand that the first 100GB is included in Preservica’s licensing fees. We also understand that we are only charged for the actual storage used. In order to compare back-end storage prices across vendors, we are choosing to price out what it would cost to dump in 1TB of data from the moment the switch is turned on and keep the whole TB in there for 365 days.**

How about verifying with Amazon that Tessella is not being overcharged! :-D Seriously, though, these are the numbers we are getting from other back-end storage providers: (We realize that Preservica as a whole is way more than just back-end storage, but we are only comparing this portion of your cost in this section...especially since it was stated that Tessella only passes the cost of the cloud storage on to its customers and that there is no mark up)

Amazon S3: $3200 for 1TB for 1 year stored in 2 geographical zones (bi-coastal)
DuraCloud: $2500 for 1TB for 1 year stored using the Amazon cloud service AND the San Diego Supercomputer Center cloud service.
MetaArchive: approx $5500 for 1TB for 1 year using a private network of geographically dispersed servers. (includes the cost of Preservation Membership, storage, and required hardware for 1 year of a 3 year contract)
Amazon Glacier: $264 for 1TB for 1 year stored in 2 geographical zones (bi-coastal) **a concern we have with this particular one is whether or not we would lose the migration capabilities Preservica offers**

How about using a different cloud provider to drive this portion of your cost down?

While our team understands that pricing is a complex beast and there are many moving parts that would be affected by even the most simplistic suggestions, we believe that there may be some viable suggestions listed that would actually make a seemingly great digital preservation solution affordable to a large market segment of mid-sized and smaller institutions.

PILOT PRICING MODEL
If Tessella believes that there is indeed a very real possibility that the general pricing model can be modified to fit the budgets and constraints of institutions with fewer resources, then we would be interested in piloting
Preservica, with its current set up/training delivery model. However, we would need to have some discussions around how Tessella is exploring the general pricing model modifications. Frankly, if we do this pilot and no changes to the general pricing model are offered by the time the Final Report is published in the late spring of 2014, then it will be a waste of resources as we would not be able to recommend the product to our target audience. It seems evident to our team, however, that the folks at Tessella have a genuine interest in the greater cause of the preservation of digital materials and that they have invested resources in helping those in our profession figure out ways to educate stakeholders in the importance of digital object preservation. That being said, we are hopeful that this interest will translate into the active, relentless pursuit of an affordable pricing model. If Tessella agrees to engage in this pursuit, then may we suggest the following terms for the pricing our pilot:

License Fees: Free testing for 12 months...no AIP's, SIP's, DIP's, etc. from this testing will be saved at the end of the pilot period
2 Day on-site set up and training at NIU: $7000
Follow up training phone calls: Not provided OR 2 free 90 minute calls
Storage: $0 We will stay within the 100GB provided by the license to help cut down on costs. To test our 1TB of data we will purge before going over the 100GB limit. However, if we go past the 100GB limit, then Tessella will pass on to us whatever costs were incurred in this overage as it would with any regular user of Preservica.

The 2 driving factors behind these proposed prices are...
1. When the grant was written and approved, Preservica didn't exist. It was therefore on no one's radar and not included as a line item in our budget....we have $0 budgeted for the testing of Preservica and can not receive additional funding for it. We are able to move a small amount of money around from other sections of our budget and believe we can afford to cover the full cost of the on-site training fees.
2. The procurement process for our institution is incredibly lengthy and cumbersome when a certain dollar threshold is passed. If we do not stay below this threshold, there is simply not enough time in the life of the project to go through this procurement process and still have adequate time to test the product and write the final report. The price we suggested above will keep us below that threshold.

I realize this is a lot to process and there are many moving parts here that must click into place for us to move forward. Please consider, Mike, and speak with your colleagues. If Tessella believes that there is simply not a feasible way to modify the general pricing model of Preservica within the time frame mentioned above, we would sincerely appreciate your honesty and frankness in telling us so. We hope, however, that we will be able to move forward together!

Again, thank you for taking the time to speak with us. Please feel free to contact me with any questions you may have.

Kind Regards,
Jaime

Jaime L Schumacher
Director, Digital POWRR Project
University Libraries
Northern Illinois University
Dekalb, IL 60115
815.753.0576
Approximately 12 months after this memo, Preservica’s pricing and delivery models were updated as shown below: [http://preservica.com/editions-pricing/](http://preservica.com/editions-pricing/) and [http://preservica.com/pricing/us-usd/](http://preservica.com/pricing/us-usd/)

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Comprehensive fully cloud hosted active preservation and public access solution</td>
<td>Comprehensive on premise active preservation and public access solution</td>
<td>Fully customizable on premise active preservation and public access solution</td>
</tr>
<tr>
<td>Ideal for small to mid-sized organizations looking for a complete out-of-the-box preservation solution that does not require any local IT resources or servers</td>
<td>Ideal for mid to large sized organizations looking for an out-of-the-box preservation solution that will be installed and managed in-house</td>
<td>Ideal for large organizations with in-house resources, looking for a preservation platform that can be fully customized for specific workflows</td>
</tr>
<tr>
<td>✅ Full suite of GARS compliant workflows</td>
<td>✅ Full suite of GARS compliant workflows</td>
<td>✅ Full suite of GARS compliant workflows</td>
</tr>
<tr>
<td>✅ Public access/discovery</td>
<td>✅ Public access/discovery</td>
<td>✅ Public access/discovery</td>
</tr>
<tr>
<td>✅ Active Preservation technology</td>
<td>✅ Active Preservation technology</td>
<td>✅ Active Preservation technology</td>
</tr>
<tr>
<td>✅ Automated DSpace ingest</td>
<td>✅ Automated DSpace ingest</td>
<td>✅ Automated DSpace ingest</td>
</tr>
<tr>
<td>✅ CALM catalogue synchronization</td>
<td>✅ CALM catalogue synchronization</td>
<td>✅ CALM catalogue synchronization</td>
</tr>
<tr>
<td>✅ Linked Data Registries</td>
<td>✅ Linked Data Registries</td>
<td>✅ Linked Data Registries</td>
</tr>
<tr>
<td>✅ Support, maintenance and upgrades</td>
<td>✅ Support, maintenance and upgrades</td>
<td>✅ Support, maintenance and upgrades</td>
</tr>
<tr>
<td>✅ One day introductory training webinar</td>
<td>✅ Single Server</td>
<td>✅ Fully customizable workflows</td>
</tr>
<tr>
<td>✅ All operations and backups</td>
<td>✅ Local disk and tape storage systems</td>
<td>✅ Multiple servers</td>
</tr>
<tr>
<td>✅ Macadaa held in fact Amazon RDS storage</td>
<td>✅ Cloud storage connectors for Amazon S3 and Glacier</td>
<td>✅ Local disk and tape storage systems</td>
</tr>
<tr>
<td>✅ Amazon S3 and/or Glacier durable Cloud storage</td>
<td>✅ Cloud storage connectors for Amazon S3 and Glacier</td>
<td>✅ Cloud storage connectors for Amazon S3 and Glacier</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Ideal for smaller archives and collections that anticipate slow growth in content.</strong></td>
<td>up to 100 GB in Amazon S3</td>
<td>$3,950 per year</td>
<td></td>
</tr>
<tr>
<td></td>
<td>up to 250 GB in Amazon S3</td>
<td>$6,950 per year</td>
<td></td>
</tr>
<tr>
<td></td>
<td>up to 500 GB in Amazon S3</td>
<td>$8,950 per year</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1 to 10 TB in Amazon S3 and/or Amazon Glacier</td>
<td>$11,950 per year</td>
<td></td>
</tr>
<tr>
<td></td>
<td>then $1,950 per TB</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>and $750 per TB Glacier</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Subsequent offerings of solutions similar to Preservica followed suit with transparent pricing and models aimed towards smaller institutions.

<table>
<thead>
<tr>
<th>Subscription Plan</th>
<th>Features</th>
<th>Annual Price</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>DSpaceDirect Small</strong></td>
<td>Annual Storage: 75 GB Standard features</td>
<td>$3750</td>
</tr>
<tr>
<td></td>
<td>The DSpaceDirect Small plan is suited for small organizations at the beginning stages of implementing a repository that have limited content and anticipate slow growth for the repository.</td>
<td></td>
</tr>
<tr>
<td><strong>DSpaceDirect Medium</strong></td>
<td>Annual Storage: 150 GB Standard features</td>
<td>$5500</td>
</tr>
<tr>
<td></td>
<td>This plan is ideal for organizations at the beginning stages of implementing a repository that have a variety of content which needs to be stored and anticipate medium growth for the repository.</td>
<td></td>
</tr>
<tr>
<td><strong>DSpaceDirect Large</strong></td>
<td>Annual Storage: 250GB Standard features</td>
<td>$8250</td>
</tr>
<tr>
<td></td>
<td>This plan is designed for organizations at any stage of implementing a repository that have a large amount of content which needs to be stored and anticipate substantial growth for the repository.</td>
<td></td>
</tr>
<tr>
<td><strong>Additional Storage</strong></td>
<td>Contact us</td>
<td></td>
</tr>
<tr>
<td></td>
<td>For storage beyond what is included in the DSpaceDirect Large subscription plan, please contact us for a custom quote.</td>
<td></td>
</tr>
</tbody>
</table>

[http://dspacedirect.org/pricing](http://dspacedirect.org/pricing)
A – 4: POWRR as a resource for instructors and practitioners - Supporting Documentation
The Digital POWRR Project became a go-to source for practitioners attempting to implement digital preservation programs and a resource for digital preservation instructors and learners.

<table>
<thead>
<tr>
<th>RESOURCE</th>
<th>PAGE VIEWS</th>
<th>URL</th>
</tr>
</thead>
<tbody>
<tr>
<td>DP101 Page</td>
<td>1,453</td>
<td><a href="http://digitalpowrr.niu.edu/digital-preservation-101/">http://digitalpowrr.niu.edu/digital-preservation-101/</a></td>
</tr>
<tr>
<td>Post-Workshop Resource Page</td>
<td>1,431</td>
<td><a href="http://digitalpowrr.niu.edu/survived-powrr-wkshp/">http://digitalpowrr.niu.edu/survived-powrr-wkshp/</a></td>
</tr>
<tr>
<td>POWRR Blog</td>
<td>1,916</td>
<td><a href="http://digitalpowrr.niu.edu/blog/">http://digitalpowrr.niu.edu/blog/</a></td>
</tr>
<tr>
<td>Tool Grid</td>
<td>10,980</td>
<td><a href="http://digitalpowrr.niu.edu/tool-grid/">http://digitalpowrr.niu.edu/tool-grid/</a></td>
</tr>
<tr>
<td>Project Wiki</td>
<td>24,847</td>
<td><a href="http://powrr-wiki.lib.niu.edu/index.php/Main_Page">http://powrr-wiki.lib.niu.edu/index.php/Main_Page</a></td>
</tr>
<tr>
<td>1-Pager Communication Documents</td>
<td>2,627</td>
<td><a href="http://powrr-wiki.lib.niu.edu/index.php/One_Pagers_tailed_to_evaluate_different_professionals">http://powrr-wiki.lib.niu.edu/index.php/One_Pagers_tailed_to_evaluate_different_professionals</a></td>
</tr>
</tbody>
</table>
The DuraSpace Community Webinar Series. Series Eight: Doing It: How Non-ARL Institutions are Managing Digital Collections Curated by Liz Bishoff, The Bishoff Group LLC
http://www.slideshare.net/DuraSpace/11122014-services-2
work on cost modeling for digital preservation and building on the recommendations of the Blue Ribbon Task Force on Sustainable Digital Preservation and Access.\(^6\)

Preserving Digital Objects with Restricted Resources (POWR)\(^6\) is another project of note. It is aiming to understand the barriers and opportunities small- and medium-sized institutions face when they have been given digital preservation responsibilities but few resources to support digital preservation activities.\(^6\) The project is evaluating digital preservation tools and services that small- and medium-sized institutions could implement. A workshop series is being planned to disseminate findings and guidance from the project. Results from these projects will help to clarify costs and improve decision-making and strategic planning, which can in turn advance knowledge about the resources needed for the long-term management and development of sustainable infrastructure for digital preservation.

In addition to cost information and models, other methods of measuring and providing evidence of the value of digital stewardship activities are needed. Improved and sharable metrics about the quality and success of digital stewardship activities can help guide decision-making. Performance statistics have long been collected and compared to help libraries evaluate and improve the management of their collections and operations.\(^6\) An effort is underway\(^6\) to determine ways of measuring collection usage across digital library platforms that inform management decision-making. This work should be broadened and built upon to give those responsible for digital stewardship the tools they need to advocate for the resources required.

Digital stewardship is important beyond the cultural heritage sector. Critical data are stewarded for scientific research authentication and repurpose; large data sets are mined for competitive advantage; and data


http://hdl.loc.gov/loc.gdc/lcpub.2013655119.1
A – 5: POWRR’s findings as reference point – Supporting Documentation

POWRR’s findings and its digital preservation functionality grid became a crucial reference point by which vendors and solution-seekers characterize and understand the capabilities of various digital preservation tool and services.

An email to a University Archivist at a small, private liberal arts college. From: Joanna Efthymiou <eftj@preservica.com>

Date: Wed, Sep 10, 2014 at 9:00 AM
Subject: Digital Preservation Report for Small & Mid-size Institutions
Appendix B – External Project Summative Evaluation

The POWRR Project
Summative Evaluation
October 2014

Amy Rudersdorf
Assistant Director for Content
Digital Public Library of America
amy@dp.la
(608) 501-1502

Review Process
The in-person review with the project manager occurred on October 23 and 24, 2014, and follow up responses from project partners were provided via email in the weeks subsequent to the in-person interview. The in-person discussions were supplemented with resources available on the website, through internal communications (email messages, draft documents, etc.), and via analytics software (for quantitative data). The reviewer has been involved as an advisor since the launch of the POWRR grant in 2012.

Initial meeting
The first meeting of advisors for the Digital POWRR grant took place on October 11-12, 2012. In attendance were representatives from 4 of the 5 partner institutions named in the initial grant, in addition to the five-member advisory committee. It was clear at that time that the experience of the group with digital preservation tools and practice was minimal; some partners expressed trepidation and others expressed outright skepticism that the goals set by the lead institution, Northern Illinois University, could be achieved. Some concern was expressed by the advisory committee, as well, that the goals were too ambitious and that the activities should be narrowed. To their credit, the NIU team felt confident that they could achieve the goals within the timeline of the grant.

Goals
The goals of the grant were related to the digital preservation needs and long-term care of digitized and born-digital content held at small and mid-sized academic institutions. These included investigating need, identifying possible tools and procedural processes, disseminating findings, and educating professionals in typically under-resourced institutions of similar size. The grant workplan was ambitious, but the main activities were (a) one-on-one interviews at each campus to better understand the current preservation-related activities being undertaken by faculty and professional staff, (b) identification of tools and services that might meet these needs, (c) consult with vendors about each of the tools and services available, (d) test and evaluate effectiveness of each of the tools and services, (e) convene a meeting of preservation professionals at the NIU campus and attend other national conferences to disseminate the findings.
Activities
The advisory board met in person on two occasions and three times by telephone over the course of the grant to help fine-tune some of the implementation plans. Some one-on-one meetings were also held with specific advisors, particularly during the development of the workshop curriculum. At the first in-person meeting, the project team leaned heavily on the experience and knowledge of the advisory committee, asking for input and advice at the very granular level and, as is more typical, support for the project’s broader goals. For example, the advisory committee played a lead role in identifying the most appropriate curation and preservation tools and service to evaluate as part of the grant. In addition, at this meeting the advisory strongly suggested that the project team reconsider the idea of hosting a conference, for the simple reality that (a) Dekalb is not an easily accessible destination and (b) the event was not one that professionals would likely choose to attend over, for example, ALA or Best Practices Exchange. The advisory board suggested that a more effective approach would be for the team to travel to established conferences to disseminate grant outcomes. Ultimately, this change in the workplan was approved by IMLS.

Some concern about the readiness of the project team was expressed after that initial meeting, but already by the second in-person meeting, the advisory board’s impression was changed completely. The work completed in 2013-2014 was impressive, and the findings were not only interesting but generally applicable to the work of preservation practitioners at academic institutions. Of particular note were three outputs completed in the first year. The first was a set of case studies undertaken by staff at each of the participating institutions to identify the state of and readiness for digital preservation activities at the various campuses, which include an urban state college, tier two state universities, and a small, private liberal arts college. The second was a gap analysis that included interviews of scholars, researchers, and professional staff at each of the institutions to gain a better understanding of the actual day-to-day digital preservation practice of and obstacles for these individuals. The third major outcome was a tool and services grid, the purpose of which is to provide trustworthy and neutral information about available digital preservation tools and services so that institutions can map the most appropriate to their existing needs, resources, and technical capabilities. In the second year, the project team continued testing and fine-tuning the tool grid, writing the white paper, and developing and hosting hands-on (and sold-out!) workshops across the US.

Feedback
The quality of work that the team has produced is exceptional, and the response from the digital preservation and stewardship community reflects that. The team—at first, at least from the view of this advisor, a coalition of semi-willing (if not earnest)—has coalesced into a community of practice that shows great signs of continuing to grow the work they have started through this grant. The tool grid, available alongside an active and informative website that includes a blog, workshop materials, and other grant outputs, has been viewed over 25,000 times. The white paper was downloaded more than 3,000 times in the first three months from Huskie Commons, the lead institution’s Digital Repository.

The tool grid and website are repeatedly referred to in educational materials and on lists of the most useful preservation resources available. The project manager and other team members have been invited presenters at a number of national and international
conferences, and have been asked to write blog posts the Library of Congress’s blog, “The Signal.” Perhaps most excitingly, their workshops continue to sell out literally overnight, and the demand for them is so great that the team has had to turn down requests (for lack of time and resources). The Society of American Archivists will audit an upcoming workshop to determine whether it could be part of their certification curriculum for digital archives specialists.

It is clear to the advisory board that their outputs were well-timed and greatly needed by the digital preservation community. A single event shows the power (POWRR!) that the momentum of this grant was able to reach. In 2013, a single email from the project manager to a service provider led to a complete revamp of that vendor’s communication and marketing policies and pricing model. Today, their offerings and costs are available to any potential buyer on their website.

**Recommendations**
The advisory committee is impressed by the breadth of work performed in such a short period of time by the project team, and feels that the participants should be proud of the good work and strong impact they have made on the digital preservation community. A few recommendations for the project team are listed below, but because of their exemplary work, these are minimal and actually fall outside of the existing scope of the grant. It is proposed that the project team pursue a second grant to undertake these activities.

- Continue its work to keep the website and related resources current, and update the tool grid as new applications and services become available or the appropriateness of existing tools diminish. If possible, the team should work closely with registries like COPTR if the POWRR tool grid no longer becomes sustainable.
- Develop workshop materials into an easily repurposable curriculum or “workshop in a box.”
- Move from the investigation phase to the implementation of tools and services at the participating institutions. Implement stop-gap measures until resources can be identified and partnerships for long-term management can be established.

It is clear by reviewing the outputs and seeing the reception that the grant team’s work has garnered, that they have exceeded the workplan laid out at the start of this project. Their work has promoted strong collaborations and developed professional relationships that will continue to engage and support the digital preservation community’s growth and knowledge. Their white paper, tool grid, and hands-on workshops go far to support the exemplary stewardship of cultural heritage collections to ensure that the discovery of knowledge continues far into the future.
Appendix C – Conference Presentations, Papers, Posters, Workshops, and Participation

Workshops

POWRR Pilot Workshop April 2014 – Chicago, IL

MAC (Midwest Archives Conference) Annual Meeting April 2014 – Kansas City, MO

NACFS (National Archive Conference for Fraternities and Sororities) June 2014 – Champaign, IL

(More information available at http://archives.library.illinois.edu/slc/research-education/schedule-class-visit/)

Data Driven: Digital Humanities in the Library Conference June 2014 – Charleston, SC

POWRR Workshop September 2014 – San Francisco, CA

AASLH Workshop September 2014 – St Paul, MN

Digital Frontiers Workshop September 2014 – Denton, TX

POWRR Workshop October 2014 – Northampton, MA

POWRR Workshop November 2014 – Evanston, IL

Presentations

ALA (American Library Association) Annual Conference 2013 – Chicago


RBMS (Rare Books and Manuscripts Section) Conference June 2013 – Minneapolis, MN

IACRL (Illinois Association of College and Research Libraries) Conference March 2014 – Chicago, IL

OAH (Organization of American Historians) Conference April 2014 – Atlanta, GA


ACRL (African American Studies Librarian Section) Research Forum June 2014 – Las Vegas, NV


ASALH (Association for the Study of African American Life and History) Annual Convention September 2014 – Memphis, TN

iPres Conference October 2014 – Melbourne, Australia (Poster available at http://commons.lib.niu.edu/handle/10843/13618)

Attended


PASIG (Preservation and Archiving Special Interest Group) Meeting 2013 – Washington D.C.

diConference (Sponsored by iSchool) Conference 2013 – Fort Worth, TX

Digital Directions July 2013 (Sponsored by the NDCC (Northeast Document Conservation Center) – Ann Arbor, MI

Digital Preservation Meeting July 2013 (Sponsored by NDSA (National Digital Stewardship Alliance)) – Alexandria, VA
GRANT PRODUCTS
February 2015

This portion of the Final Report packet contains Grant Products that are referenced in the Narrative section of the Final Report packet. Included are the following:

1. Data Accessioner information
2. POWRR Tool Grid information
3. COPTR/POWRR partnership
4. The POWWR website, blog, and resources
5. All Partner Organization case studies
   a. Chicago State University
   b. Illinois State University
   c. Illinois Wesleyan University
   d. Northern Illinois University
   e. Western Illinois University
6. POWRR White Paper
7. POWRR Workshop Documents
   a. Expected Outcomes and Prerequisite sheet
   b. Handouts
   c. Slides
8. Communication 1-pagers
9. Internet Archive Tutorial
10. Intellectual Capital at Risk: Data management practices and data loss by faculty members at five American universities
11. Results from interviews at all Partner Organizations
12. Analysis and model agreement for developing a cooperative network for distributed digital preservation
Data Accessioner

The free, open source, user-friendly digital collection “triage” tool that was updated and enhanced through the POWRR project and is used in POWRR workshops for hands-on learning.

Data Accessioner

Download binaries for all versions on the "Version Notes & Binaries" page.

About

The Data Accessioner (DA) is a simple tool, with an easy-to-use graphic interface, for migrating content between media while also:

- creating and validating checksums,
- gathering metadata (via FITS),
- and compiling an XML metadata file (with the option to include Dublin Core metadata as of v 1.0) for future reference.

The Data Accessioner was built out of the need for a simple GUI interface to allow the Duke University Rare Book, Manuscripts, & Special Collections Library (now the David M. Rubenstein Rare Book & Manuscript Library) Technical Services staff an easy way of migrating data off disks and onto a file server for basic preservation, further appraisal, arrangement, & description. It also provides a way to integrate common metadata tools at the time of migration rather than after the fact. With a simplified interface and being written in Java it is intended to be easily adopted by smaller institutions with little or no IT staff support.

The very first version of the tool was written in the course of a week in early 2008 and, although usable, it was more of a proof-of-concept. In January 2009 the Data Accessioner was revisited with a revised architecture. Also, the metadata tool adapters and the custom metadata manager where extracted to be used as plugins.

The most recent versions (0.3.1 and 1.0.0-beta) were made possible by the POWRR Project. At of these versions the Data Accessioner is no longer formally associated with the Duke University Archives or their parent organization.

Find the source code on GitHub.

New! For those of you who are uncomfortable reading (or transforming) the XML-based metadata files produced by the DA I have created a new tool to help. Take a look at the Data Accessioner Metadata Transformer.

Updates don’t come very often but if you would like to now when something happens you can add the “Updates” page to your RSS reader or just check back from time to time.

Feedback and questions are always welcome via email!
Tool Grid

The team created a DP tool evaluation rubric based on the OAIS digital curation and preservation lifecycle, and conducted an environmental scan of DP tools, technologies, and services available as options to institutions with fewer resources. The master list of options and the rubric can be found on the wiki at [http://powrr-wiki.lib.niu.edu/index.php/Tool_Functionality_Evaluation](http://powrr-wiki.lib.niu.edu/index.php/Tool_Functionality_Evaluation). For the 60+ tools, services, etc. that were not selected for in-depth testing, the project team used this rubric to create the POWRR Tool Grid, a graphic-based resource for practitioners that shows, at-a-glance, the functionalities of the DP tools and services and how they fit within an OAIS-based digital curation lifecycle.

![Tool Grid](http://digitalpowrr.niu.edu/tool-grid/)

URL: [http://digitalpowrr.niu.edu/tool-grid/](http://digitalpowrr.niu.edu/tool-grid/)
COPTR Partnership

As use of the POWRR Tool Grid increased and became more widespread, the team was contacted by representatives of an organization in the UK working to consolidate the many DP tool registries available on the Web into a single registry, COPTR, a Community Owned digital Preservation Tool Registry. Knowing that the Tool Grid was merely a snapshot in time and its upkeep not sustainable by the grant, the team chose, along with several notable institutions, to support this effort by providing them with an export of the Tool Grid data. Given the popularity of the graphic-based look and feel of the Tool Grid, the POWRR team commissioned the steward of COPTR to combine the form and function of the original POWRR Tool Grid with the far greater coverage of tools and sustainability provided by the COPTR data feed, resulting in the POWRR Tool Grid v2. http://www.digipres.org/tools/about/

About the POWRR Tool Grid

The POWRR ToolGrid v2 provides a set of interactive views designed to help practitioners identify and select tools that they need to solve digital preservation challenges. Everything in the Grid is hyperlinked, so simply click through the displays until you find the information you are looking for. Clicking on the name of a specific preservation tool will reveal more detail on the COPTR wiki. As well as the main Tool Grid that makes navigation more manageable there is also an Unfiltered Grid that lists every tool in every content category against every function.

This Grid is based on the Tool Grid first developed by the Digital Power Project, and combines the form and function of the original POWRR grid with the far greater coverage of tools provided by the COPTR data feed.

How does the POWRR Tool Grid work?

The POWRR Tool Grid is re-generated nightly from the contents of the COPTR wiki, if you have any changes or additions to the data in the Grid, sign up for the COPTR wiki and make your contributions there.

Feedback

We very much welcome feedback on the POWRR Tool Grid. Please contact ..., or discuss any Grid issues at the international collaboration about digital preservation Google Group.

How do I contribute new information to the tool grid?

Easy. Sign up for the COPTR wiki and make your contributions there. They will propagate to the Tool Grid overnight.

Attribution

All data presented in the Grid is sourced from the COPTR wiki under a CC BY-SA licence.

Funding and development

Development of the POWRR Tool Grid was made possible by the Digital Power Project and the University of Kent. Development work was completed as part of a collaboration between the Digital Power Project and the COPTR consortium.

URL: http://www.digipres.org/tools/about/
Main Page

Community Owned digital Preservation Tool Registry (COPTR)

COPTR describes tools useful for long-term digital preservation and acts primarily as a finding and evaluation tool to help practitioners find the tools they need to preserve digital data. COPTR is balancing the knowledge of the digital preservation community on preservation tools in one place. Instead of organizations competing against each other with their own registries, COPTIR is bringing them together. In doing so COPTIR’s objective is to provide the best service for practitioners on digital preservation tools.

There are 391 different tools described in COPTIR.

Browse the COPTIR Registry

- Find the tools you need in the all-encompassing Tool Grid, which is POWERed by data from COPTIR.
- View all tools
- View tools by functional category
- View tools by the type of content they act upon

How to create a new Tool Entry

1. Check out the Guidelines for contributing to COPTIR.
2. Search for the tool you want. Check the full name and the summary.
3. If you find it, consider adding more detail to the adding entry. If you don’t find it, follow the link to create a new entry, which will automatically create a outline entry for your tool based on the Tool Page Template

COPTIR Partners

COPTIR was created and launched with the support of the Adapting National Approaches to Digital Preservation initiative and was initially populated with data from registries run by the COPTIR partner organisations.

URL: http://coptr.digipres.org/Main_Page
The POWRR website, blog, and resources

Webinar Highlighting the Digital POWRR White Paper

The POWRR White Paper was highlighted during a recent DuraSpace Hot Topic webinar. One of the presenters was Liz Bishoff, the principal partner of the Bishop Group LLC. Liz is also a member of the POWRR Advisory Board. DuraSpace offers several excellent webinars free of charge on topics ranging from Digital Preservation Planning to Research Data Management Support to Managing and Preserving Audio and Video in your Digital Repository. The POWRR team found the folks at DuraSpace to be very responsive and helpful as we conducted our research and piloted their cloud-based preservation service, DuraCloud.

Discussion on the Outcomes of the Digital POWRR Project

Take a look at this opportunity to hear members of the POWRR Team discuss the trials, tribulations, victories, and the future of the Digital POWRR Project.

"The NDSA Infrastructure working group invites you and your colleagues to a call on the outcomes of the Digital POWRR project. In keeping with our ongoing series of conversations, you can expect about half the call to be a presentation and the other half to be time for conversation and discussion.

Title: The Digital POWRR Project: What we discovered. what we did about it, and what still needs to be done.

Abstract: Lynne M Thomas and Jaime Schumacher will discuss the outcomes of this IMLS National Leadership Grant project, outline those deliverables that were particularly well-received by the community, identify gaps that have yet to be addressed, and, with the project end-date approaching, seek guidance on the transfer of project-created products that should be maintained and cultivated for the benefit of the wider community.

When: November 18, 2014 at 2pm ET
Call in #: 877-299-5123*

Preservation processing update

DataAccessioner developer Seth Shaw just sent a tool to help with reports analysis. He says he wants to get feedback on a simple report transformation (from .xml to .csv) tool first. After that, he's going to add a way to aggregate the data from the .csv into size by type of file, etc. within the DataAccessioner.

He's created a DA-branded version of his XSLTProcessor and named it the DA Metadata Transformer (DA-MT); see image below. You can download it at http://dataaccessioner.org/downloads/da-mt/da-mt.zip

URL: http://digitalpowrr.niu.edu/blog/
New to Digital Preservation (DP)?

Basics and Background

- Wikipedia Overview
- Digital Preservation Europe – Visit their YouTube channel for cartoons about digital preservation.
- Preserving Your Personal Digital Memories (Webinar)
- DPE Digital Preservation Video Training Course
- Digital Preservation in a Box
- Rescuing the Tangible from the Intangible
- Digital Preservation FAQ - The National Archives information about Digital Preservation
- Personal Preservation Tips – More links to information about personal preservation

Practical and/or Incremental Steps for DP

- I'm new to Digital Preservation – Information from the North Carolina Department of Cultural Resources
- You've Got to Walk Before You Can Run: First Steps for Managing Born-Digital Content Received on Physical Media – Ricky Evan's article about getting digital born material off of various physical media, A step by step guide.
- Resources for Technical Steps – Additional resources to get running **See above “You've Got to Walk Before You Can Run” as well**
- Starting Small: Practical First Steps in Digital Preservation
- File Naming Conventions – “From Page 2 Post” blog post about file naming conventions
- Implement a TDR – A Blog Post by Chris Prom
- Simple E-records Preservation and Access Plan – A Blog Post by Chris Prom
- Commit to an Information Network – Digital Preservation Toolkit

URL: http://digitalpowrr.niu.edu/digital-preservation-101/
I Survived a POWRR Workshop: http://digitalpowrr.niu.edu/survived-powrr-wkshp/

Which articles were required reading for the workshop?

- Wax this Way: Detailed Steps for Transferring Born-Digital Content from Media You Can Read In-house
- You've got to Wax Before You Can Run: First Steps for Managing Born-Digital Content Received on Physical Media

I'd love to have a copy of the slides they used!

- Here it is: POWRR Workshop Slides
- A more condensed version of the slides used at NAGPS available here: NAGPS Slides

Where can I get the accessioning tool... Data Accessioner?

- Look!
- Data Accessioner has been undergoing changes! Find out all about it here! You can also keep up with the latest updates here!
- Click here to find out more and download Data Accessioner!

May I have a copy of the simplified inventory Excel spreadsheet we used for the accessioning exercise?

- Sure! You can find it here: Digital Collections Inventory

Where can I find more information about the tools and services POWRR tested and talked about during the workshop?

- Archivematica
- Curator's Workbench
- DuraCloud
- Internet Archive
- MetaArchive
- Preservica

They mentioned something about having made a tutorial for very small memory institutions (like volunteer-driven historical societies) to upload content to Internet Archive for free?

- Sure! Here it is: A guide to using Internet Archive: Internet Archive Tutorial

I loved that NDSA Levels of Preservation grid... where does that live?

- Click here to learn more about the NDSA Levels of Preservation or download a PDF version: NDSA Levels of Preservation grid.

Do you have a Digital Preservation Decision Flowchart?

- Yep! Brought to you by the University of Utah: Univ Utah DigPres Dec Flowchart

Where can I get a copy of the 3-3-3 Action Plan activity sheet they used in the workshop?

- Here you go!
CSU Case Study

Final Draft February 2014

Institutional Bio and Back Story

Metrics -Institutional Level

<table>
<thead>
<tr>
<th>Number of Students and faculty</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Students-6,107*</td>
<td>Faculty and Staff</td>
</tr>
<tr>
<td>Undergraduate-4,618*</td>
<td>Full-Time-297*</td>
</tr>
<tr>
<td>Graduate-1,489*</td>
<td>*based on Fall 2012</td>
</tr>
</tbody>
</table>

Endowment and budget

Endowment

~3 million

Budget

Operating Budget

Library Acquisitions Budget (print, e-books, and media): $938,745.19 (based on FY2013)

Carnegie Classification

- Undergraduate Instructional Program: BaI/SGC
- Graduate Instructional Program: S-Doc/Ed
- Enrollment Profile: HU
- Undergraduate Profile: MFT4/I
- Size and Setting: M4/NR
- Basic: Master's L

Colleges/Academic Programs Offered

- Degree-granting colleges
  - Arts and Sciences
  - Business, Division of Continuing Education and Non Traditional Programs Education, Health Sciences, Pharmacy
- Degrees offered
Metrics/Library Level

Size of Library
- Number of Faculty and Staff
  - 46 total FT Staff (FY 2012)
  - 24 librarians and professional staff (admin)
- Total number of volumes
  - 475,678 in Main Library
- Size of Digital Collections
  - These numbers are estimated
    - Digital Collections
    - Archives: ~1.5 TB (photographs, theses, newspapers, audio visual)-located on external drives and digital repository

Staff
- Number of staff currently devoted to digital preservation-4 (these are not full time roles)
- Organizational role
  - Haykal, Aaisha-University Archivist
  - Kong, Martin-Systems Librarian
  - Porter, Gayle-Special Format Cataloging Librarian
  - Mathias, Henty-Digital Initiatives and Web Services
- Responsibilities beyond digital preservation
  - The above positions/people have other responsibilities including administrative duties, processing collections, public programming, management of the web servers, management of the library technology, cataloging new acquisitions (media and e-books), collection development, coordination of electronic thesis and dissertations submissions.

Budget
- The Library's budget for digital preservation comes from human resources/skills and the library's annual contribution to Consortium of Academic and Research Libraries in Illinois (CARLI) as payment for the ContentDM services

Digital Preservation Technologies Currently in Use
- Eloquent Archives
  - Houses collection information including information regarding metadata and location information for collection. Additionally, this is the public catalog where we input/store records describing our archival collections. The catalog includes some digital files hosted therein. The system is hosted internally.
- ContentDM
o This is a content management system hosted externally by CARLI. The majority of our digital files (pictures) can be located here. The system is hosted externally.

- **ProQuest ETD**
  o A new process we have in place at CSU in order to provide access to our student's thesis and dissertations. Students have the ability to embargo access to their publication, but it is still stored on ProQuest servers (and a local copy is kept).

- **Internet Archives**
  o This system is linked to ContentDM. Some of our publications (yearbooks and historical books) can be found through here. A copy of the documents is also housed locally.

**Equipment**
To create the digital content within the library we are using the following BookEye Scanner (KIC) for the print; Microfilm ScanPro for microfilm digitizing; Epson Expression 11000XL for photos, and a Canon digital camera. The Bookeye is about 2yrs old. The Library Technology department is responsible for maintaining the equipment.

**Defining Moment**
The realization that digital preservation was important to us as an institution occurred when we began creating digital content and obtaining grants that paid for digitization, but not for preservation and/or storage. Furthermore, with library staffing changes, at times the only copies of digital files are currently being stored on jump drives, external drives, some on library servers, and hard drives. There lacks a clear workflow in terms of where everything is currently stored, who is responsible for what, and what goes where.

**Self Assessment Results**
After taking a self assessment survey it is possible to see that the Chicago State University Library and the CSU Archives and Special Collections has done much to get digital initiatives created and providing access to these records. Collaborating with the CSU Library allows CSU Archive material to be accessible; however, our institution lacks resources to support digital preservation of these materials. For example, CSU Library staff limited time for digital preservation work due to commitments and previously assigned duties, our digital program would be able to flourish if we had a full-time staff committed to creating and managing the digital items we produce in the Archives. Additionally, more hands-on training for staff and faculty on the issues of metadata and digital preservation would make our institution well-rounded. The assessment also made it clear to us that the funding we receive for digitization and digital preservation comes from soft-money (grants) and we need to allocate more appropriated dollars/budget toward this work, and resources could come in many forms including staff and funds to purchase the equipment and services.
The assessment also indicated to us that we need to be thinking more about our collections in terms of risk, specifically in relation to our audiovisual collections that are in our university collections and in the special collections. We have a project plan underway to convert the historically important video and audio material to a digital format in 2014-2015 so we will not lose this information. However, one of the issues that arose in the assessment is that we need have to have a policy in which we specify what types of formats we will and will not maintain and which ones we are not; and how long we will maintain these resources.

One of the first steps we plan to take is to ensure that digital preservation is written into the Archives and Special Collections mission and goal statements. Then we need to set up a clean workstation (one in which has limited software and documents) that will serve as our digital acquisition workstation.

**Policy Gap Analysis**

**Where we are**

There are three major departments that are involved with digital preservation/initiatives. They are University Archives and Special Collections, Technical Services, and Public Services/Reference. The University Archives contains historically relevant records about the campus. The Special Collections consists of donated papers from historically and culturally relevant people and organizations of Chicago’s South Side and African American communities. Technical Services inserts the metadata for the digital collections and maintains the technical infrastructure for the library. Public Services/Reference can initiate/propose items from the collection to digitize.

**Digitization**

The digitization of our archival and special collections has been possible through grant funded projects. These digital objects are accessible through ContentDM software hosted by Consortium of Academic and Research Libraries in Illinois (CARLI). In addition, because CARLI mandates that institutions store a digital copy, we have one on our server. Due to staff changes the uploading of additional content has been limited, thus, we have been saving our digitized items in-house on external drives as well as uploading them to the archival catalog-Eloquent.

**Records Management**

Since the University Archives oversees the records management program, we are responsible for the electronic records for the campus. We have recently begun exploring the usage of Eloquent Records as an electronic management software option. In addition, the Archive department is working with the Legal department and Information Technology office to implement policies and procedures for university records management as well as a method to capture and preserve the historically relevant information.
Workflow
When we receive/acquire digital materials, we document the number of items (ex. 32 CDs, 1 floppy disk, etc.) and whether a label exists. However, we lack a systematic workflow or method of viewing each CD to do an inventory of its contents and to check for possible corruption. All of this is done on staff work computers, as we lack a standalone computer dedicated to these tasks.

Where we want to be
Ideally, CSU would develop/select one set of guidelines, workflows, and storage for all of its digital objects designated for long term retention (digitization, born digital materials, e-manuscripts, university records). We need a digital collection development policy that sets priorities for what gets digitized, who manages the materials, etc. We also need to further develop campus relationships that allow us to leverage what we know about the library’s needs to help meet the needs of the entire campus.

We also would like to have either a product or service that would allow us to preserve and manage our digital collections at a relatively low cost; however, the administration would be willing to pay for a product/service that provides a robust system that serves most of our needs. This can either come at an institutional level or within a collaborative model.

The gap
The gap includes staffing, money, and education. The current staff is already busy and committed to other projects. Since staff job descriptions lack mention of digital preservation duties, it is difficult to assign staff to work on other projects. However, we do have a central staff who are educated about said duties, have expertise therein, and want to be engaged, but they too are overworked. Therefore, having money to hire an employee who is solely responsible for managing and coordinating with digital projects would be of great assistance. Barring an influx of funds, we will have to strongly encourage the administration to repurpose/retrain staff to fulfill this role. The IMLS grant has indicated to us that digital preservation in any combination of tools will have to come at a cost and thus, we have to repurpose already limited funds to this issue. Library IT is a separate unit from campus IT in our institution, however, the campus IT has been supportive of what the library aims to do, as long as it does not conflict with their plans for the university network structure. The relationship between the two units is cordial. However, it is important for the language/terminology used to be interoperable and understandable for each area.

Our campus is supportive of collaborative models, so having a solution that includes this would be suitable for us. The library staff currently has various levels of understanding of what digital preservation entails; thus, it would be ideal if those who have had the opportunity to learn more about it would share their knowledge.
The recent Illinois SB 1900 open-access bill mandates that “each public university shall establish an Open Access to Research Task Force” by January 2014. Therefore, it is essential for each public university to commit to “preserving the archival version of the resource in its highest quality in accordance to collection development policy retention decisions. The bill has opened the conversation on our campus, and now that there is an impetus on this issue we can most likely receive financial support to address it and implement its practice.

Digital Preservation Policy/Program Proposal

The policy/program proposal is in draft status and the final version will be ultimately decided by the CSU Library administrative body.

Mission and Scope

Chicago State Library Digital Preservation Policy supports the preservation of digital resources that are included in the Library's collections. These digital resources are subject to the same criteria for selection and preservation as other resources in the CSU Libraries. These decisions regarding selections and preservation are made by selectors, curators, and bibliographers as experts on the value of the content, in consultation with the relevant information technology and preservation experts. Digital preservation decisions are made on the basis of this Policy, the Library's Strategic Plan, the digital resources' enduring value and the feasibility of the digital resources' preservation. When possible, decisions about the need for preservation are made at the time of creation, acquisition, or licensing of digital resources.

Selectors, curators, and bibliographers in consultation with technical experts must specify the preservation requirements for the digital resource. Preservation responsibility is retained by CSU whether the digital resource is preserved at CSU or entrusted to an outside agency. Preservation of digital resources may include any actions necessary to preserve continued access to the digital material, ensure its authenticity, mitigate and/or reverse the effects of hardware and software obsolescence and media decay.

This Policy recognizes that the maintenance and the reliable long-term access to CSU’s digital resources are supported by a preservation planning function. A core activity to preservation planning includes research on and planning of technology that supports a repository and the requirements of the designated community. Other important activities of preservation planning include outreach and education regarding policies, procedures and best practices for digital resources.

Digital Preservation Workflow

Selection

Selection of digital files to preserve will be performed by the appropriate subject specialists and/or unit and department heads in accordance with established collection development policies for those collections and units.
Acquisition

Initial acquisition of relevant born-digital files will be performed by the appropriate subject specialists and/or unit and department heads in accordance with established collection development policies for those collections and units. Materials will come from donations and from university offices/departments/units record transfers. Additionally, small volume scanning of paper and photographic objects can be performed in house. Scanning of large volume items and other collections (audio, video, etc.) will need to be outsourced. Individual units and content creators are responsible for overseeing this project and for furnishing relevant files in appropriate formats for preservation to the Digital Initiatives staff for long-term preservation.

Content Types

Each of the above content sources mentioned may present content in one or many of the following types, which may require different preservation strategies due to their varying attributes.

- Textual materials (ebooks, articles, etc.; ASCII, UTF-8, Unicode)
- Images (scanned books or photographs, digital photographs, digital art; TIFF, JPEG, GIF, JPEG2000)
- Audio/video materials (videos produced on campus, recorded sound oral histories, etc.; MPEG, AVI, MOV, AAC, WAV)
- Numerical data/datasets (research data; XML, XLS, proprietary database formats)

The library will likely acquire materials in additional formats in the future, and preservation strategies will be developed to accommodate new formats as needed.

Curation

(Managing files and metadata): File management will be completed in the selected system, governed by collections-based policies for availability, access, and metadata creation. Metadata will adhere to professional standards (i.e. MARC, MODS, EAD etc.). The goal is to have single entry of metadata promulgated across multiple systems if necessary to promote metadata promiscuity. Processing of digital materials before submission for AIP conversion is the responsibility of the collection managers (e.g. SIP creation/Archon record creation/finding aids/MARC records where applicable). Preservation metadata will be system-generated (i.e. Archivematica/Curator’s workbench), based on policies set in coordination with collection managers, technical services staff, and Digital Initiatives staff.

Archiving

The University Archivist, Special Formats Cataloger, Systems Librarian, and the Digital Initiatives and Web Services staff will be responsible for creating and managing Archival Information Packets (AIP) in preparation for storage, in consultation with individual collection managers to ensure appropriate metadata assignment and access levels, providing packet-level metadata, and uploading packets to the selected storage solution.
Storage
Storage needs to be locally redundant and geographically redundant/dispersed; accessible only to relevant staff only; live networked storage (rather than static CDs/DVDS/Hard drives/external drives); scalable; and affordable. We are still negotiating how library storage and campus storage will intersect; we expect cooperation between campus IT and the library to meet our needs.

Retrieval
Creation of Dissemination Information Packet (DIP): will be dictated by collection managers in terms of formats, etc. Actual creation of DIP will be handled by Digital Initiatives Staff on the fly, by the repository system or by patron request. Creation of DIPs may vary based on whether the DIP is designed to be publicly available or furnished for research purposes only (i.e. not public). Theoretically, patrons would be able to request specific folders of records/documents after viewing an Eloquent finding aid. The corresponding records/documents would not be publicly attached to the finding aid in collections whose materials are not in the public domain. A separate site created for serving access copies of public domain materials would allow patrons to download them directly, without having to make use of the DIP system.

De-selection
Digital objects will be reviewed and disposed of as needed, based on collection development policies.

Strategy

Communication and Education

- **Administrators**: Senate Bill 1900 encourages resource allocation for open access and long term storage. Administrators need to be educated regarding the importance of said bill and its ramifications in order to make funding commitments. Education work can be done through one-on-one meetings, workshops, and presentations.

- **Colleagues who are stakeholders**: Office of Sponsored Projects, Faculty (need more outreach), Library subject specialists (more outreach), Graduate School (ETDs), and Campus Information Technology Department

- **Content providers**: University Archives, Record Management, Special Collections, Digital Theses and Dissertations, Library Faculty (Digital Projects), Office of Scholarly Communications, University Units/Departments/Offices, and faculty (open access publications)

- **General marketing/PR**: University and Library Website. Person-to-person contact to faculty and academic departments through library liaisons. Developing and hosting a symposia/conference to garner awareness (authors rights/open access/long-term preservation). Create brochures and marketing
material about open access and digital preservation. Work with library and university PR offices to create a message.

Allocation of resources


- Budget: This will be determined based on salary for above positions and funding for a service and/or product. We will need budget for: On-campus redundant servers. Dedicated hard-funded allocation for digital preservation activities. Regular budget for local servers/storage space and redundant campus backup. Budget for subscription fees for geographically disparate offsite digital storage.

Policy Review

This policy and the actions that flow from it will be evaluated regularly to ensure that implemented strategies continue to support the Library's mission and policies, use resources in a cost-effective manner, and adapt appropriately to address evolving technologies. This evaluation will be completed at least once every three years.

Content Types

Each of the above content sources may present content in one or many of the following types, which may require different preservation strategies due to their varying attributes.

Lessons Learned

Going through the testing of the various tools and services and the creation of a draft policy for Chicago State University Library and Archives has let us know that we have come a long way and that we have staff that are willing and have the skills to assist us in this process of digital preservation. However, it has also shown us that we have to educate colleagues in the library and the campus as a whole due to misunderstandings about what digital preservation is about and how one needs to ensure long term access to the digital items we are creating on a daily basis. This is clearly seen in the on-campus interviews that were undertaken where some faculty are doing nothing to protect their data and others are consistently storing their data.

The creation of a digital storage environment is going to involve many constituents (known and unknown at this time), therefore any product or serves that we use is going to have to be tiered as it not cost-effective or recommended to preserve all content the same way or for the same duration.
Evaluation of Tools/Services Tested-Reflections and Recommendations

Test Files
We tested files from our special collections, university archives, university faculty and ETDs. Most of these files did not have associated metadata, so we created Dublin Core metadata for the files.

Formats
DOC, JPEG/JPG, MXD, PDF, PNGs, PPT, PPTX, SAV, SPSS, TIFF, WMA, WPO

Recommendation Levels
Highly Recommend    Moderately Recommend    Do Not Recommend

DuraCloud

Reflection
We like the fact that DuraCloud service provides two geographic separate locations to house our data. It is a cloud based solution, and thus requires nothing to be locally installed to get it up and running. However, if a user wanted let it sync to a network location, the sync tool would have to be downloaded and locally installed. Other features that was appreciated was the ability to add tags to the files once it is in the system as well as the pie graph that let us know the types of files we have in the tool as well as the formats of them.

DuraCloud sync tool had the option of “sync deletes” which when enabled, evidently causes mass deletion of files if the original directory and file is not found by the sync tool on our end (ex. If two institutions are sharing a DuraCloud space, when institution X does a sync, Institution Y’s files are deleted due to the fact Institution X does have Y’s files to “sync”). This lends itself as a possible problem if multiple partners and/or institutions are working together in one “space” in DuraCloud, whereby one individual can inadvertently delete the files of another individual simply by choosing the option of syncing deletes when running the sync tool against the DuraCloud server.

Recommendation- Moderately Recommend
DuraCloud would be an immediate solution to the digital preservation problem for a small institution as it does not require a lot of time to manage or staff for support. In addition, one has easy access to the files and can download from the system. For CSU we would use it for several formats including, pdfs, office documents, and photos.

Curator’s Workbench

Reflection
The product does not provide a manual software installation, but does include a video on it and help section within the tool. In addition, one has to make sure the workstation has the right version of Java installed and also have software to extract files. Curator’s Workbench is a processing tool that creates a METS wrapper with MODS elements, however, if the user is not familiar with these standards the product is difficult to use. One of the services that the tool provides is a crosswalk between the metadata that you have created and MODS, but if your data
is not in the right sequence/order then the crosswalk does not work. Additionally, if your collection is complex it can take a lot of time to create the crosswalk.

After limiting our elements to a few fields a METS wrapper was created, which one could then export into another system. A nice feature of the system is that it creates a bag that one could then use to input files into a system such as MetaArchive. When we corresponded with the developer he indicated that although the current Bagger version used by Curator’s Workbench is not updated, UNC plans to do so.

**Recommendation - Do Not Recommend**
CSU would only use this tool if we used MODS metadata standards to markup our collections, however, we use Dublin Core and EAD standards, thus, we do not recommend this for our workflow.

**MetaArchive**

**Reflection**
The MetaArchive service used the LOC bagger tool to bag the files that an institution wants to preserve/harvest on the MetaArchive servers. The installation of Bagger was relatively easy; however, one has to have an understand how to extract files from a zip drive as well as the proper program on their workstation. Bagger allows the user to insert a file with one’s metadata on it for each file and then create a wrapper with metadata for the whole package. After trial and error due to following incorrect instructions, we were able to create a template for the wrapper, which makes the product more customizable. However, we did wish that there was a way to directly associate the metadata with the file. Creating the bag to send to MetaArchive involved us sending our files to Northern Illinois University (NIU), which in our trial served as the “node,” which MetaArchive’s servers would then harvest our files from. We reached an impasse when FTPing our files to NIU’s server from our campus due to firewall and ISP issues. Nevertheless, working with our campus IT staff, we were able to transfer the files from our campus to NIU. After the transfer, our work was done and we just had to wait to see if MetaArchive could harvest our files. MetaArchive is file agnostic, meaning it does not care what type of files you have; it harvests all formats. However, once the harvest was complete we found out we had three unusual files from our faculty that could not be harvested due to server permissions at NIU.

It is important to note that finding the issue with customizing the metadata did take the expertise of our systems librarian and some web searches to locate the issue. However, due to our trials, the instructions have been updated to reflect the correct information.

The customer service representative worked with us through the entire project and was responsive to e-mail requests.

**Recommendation - Moderately Recommended**
This service would be good for us at CSU because it has geographically displaced nodes for our material and we would be easily recoverable in case of a natural disaster. The limit for MetaArchive for us as an institution would be that we could not support being a node by ourselves; we would have to go in as a collaborative/cooperative membership with other
organizations. In addition, one has to request migration of records to different formats and it does not provide an access portion to the files, it is purely a “dark” archive that preserves the bits of files.

Archivematica

Reflection
Throughout the installation and testing phase of the grant Archivematica moved through three different versions (0.9-beta, 0.10-beta, and 1.0). The major issue about Archivematica is that one has to install a virtual machine in order to run the software or acquire another machine that you can install Linux on it, and there are instructions for it on the wiki page. The relatively more difficult part of the installation was extracting the Archivematica Zip file due to unfamiliarity with the tgz file type. Once this is complete the user has to set up Secure Shell so that data files can be transferred to the virtual machine. It is important to note the difference between host-only and bridged networking if the institution wants to be able to access the machine from a different workstation; such configuration changes must be made in the /etc/network/interfaces file on Linux in order for it to work.

Before testing the software, we watched the video that is on the wiki page, which was useful in helping us determine the process. As the user goes through the product, it asks you to confirm several different steps (one can select a default choice through the administrative settings to facilitate future processing). Also it tells a user when something is wrong/when it fails and why it failed (and sends the admin an e-mail), but it does not tell you how to fix it in the cases we had. Some of the issues we had included having PDF’s not fully formed correctly and having our system time out so that the AIP could not store properly.

One of the system features is ICA-Atom, which the institution could use as their content management system as the DIP that is created can be deposited here. If one does not want to use ICA-Atom, then they could use their content management system. We use ContentDM, but we did not test the two together. We figured out that if files are not recognized by Archivematica, then they do not show up in ICA-Atom. One of the major problems we had was that we could not get our imported metadata to be shown in the DIP or under the AIP when there was more than one file being processed.

We made heavy use of the user discussion group in Google Group during our testing phase of the product; however, we did not get any responses from Archivematica users, but from Artefactual (the developers). Artefactual representatives were very patient and assisted us with our issues. One of the responses to our questions led us to install version 1.0. As of February 2014, there is no virtual box image for Archivematica so one has to install version 1.0 via command line from the packages. This can be daunting for someone who does not have Linux command line expertise. The library’s systems librarian did this over a period of 12.25 hours.

Recommendation-Highly Recommended
If CSU were to choose Archivematica as a solution we would have to work out the metadata importation issue, which would probably involve hiring Artefactual for consulting hours on this
issue and any future ones we have. Additionally, before any full implementation of the product we would have to see how it works with ContentDM.

Preservica

Reflection
The Digital POWRR team had a two day demo via webinar and hands on training of the Preservica software by the Preservica UK team. The demo walked through several various processes (workflows) that the system can do including web crawls and migration of data. The workflow follows the Open Archival Information System (OAIS) model very closely and the user can see where in the system they are. The system uses Amazon cloud based web server for the storage of your data. Additionally, the product provides active preservation steps where the user can set up what occurs to each file and there is documentation to see what occurred and when it happened. The product interface is intuitive and tells a user what to do when and gives users options in regards to several steps including if they only want to create a preservation copy or an access copy or both.

We at CSU also appreciated the fact that it has an access portion that one can use as the content management portion of the workflow and that there are permission limits for what a user can or cannot do; this is especially important because there could be many staff involved. Furthermore, the Preservica team is continuing to update and refine the product to meet user needs and concerns. An example of this is the new collaboration between them and the developers of DSpace, so institutions could import directly into Preservica without having to do a two step process. Unfortunately, no collaboration with OCLC for ContentDM is in the works, but possibly the more US users they get this will be an option.

Preservica was one of the only systems that we tested/used that ingested completed metadata easily and without many hurdles, which is a big plus for us, so one does not have to do double the work.

Recommendation-Highly Recommended
This is a full service product. It provides processing, preservation, and access to collections. A limiter for CSU is the price, however, we have been informed that in the coming year a new pricing structure will be developed.

Appendix 1: Standards and Best Practices

Chicago State Library will observe national and international standards and best practices for the creation and management of digital objects, along with the associated metadata needed to maintain resources throughout their lifecycle. Open source formats will be preferred.

Relevant standards include:

- Open Archival Information System Reference Model (OAIS)
- PREMIS Data Dictionary for Preservation Metadata
- Trustworthy Repositories Audit & Certification (TRAC): Criteria
ISU Case Study

April 2014

I. Institution Bio & Back Story
A. Metrics – Institutional Level
1. Number of Students and Faculty (Fall 2012)
   - On-Campus Total Enrollment: 20,502
   - Undergraduate Students: 18,207,
   - Graduate Students: 2,295
   - 3,563 University Employees
   - 1,212 Departmental Faculty (239 Professors; 242 Associate Professors; 218 Assistant Professors; 513 Other),
   - 41 Non-departmental Faculty; 29 Library Faculty,
   - 112 Laboratory School Associates,
   - 760 Administrative/Professional,
   - 1,409 Civil Service,
   - 97.3% of Tenure/Tenure Track Faculty hold a terminal degree

2. Endowment and Budget
   2011 endowment, $69,194,754

3. Carnegie Classification
   - Undergraduate Instructional Program: Prof + A&S/HGC: Professions plus arts & sciences, high graduate coexistence,
   - Graduate Instructional Program: Doc/Prof: Doctoral, professional dominant,
   - Enrollment Profile: VHU: Very high undergraduate,
   - Undergraduate Profile: FT4/MS/HTI: Full-time four-year, more selective, higher transfer-in,
   - Size and Setting: L4/R: Large four-year, primarily residential,
   - Basic: DRU: Doctoral/Research Universities

4. Colleges/Academic Programs Offered
   Degree-granting colleges
   - Applied Science and Technology,
   - Arts and Sciences, Business,
   - Education,
   - Fine Arts,
   - Mennonite College of Nursing,
   - Other (Interdisciplinary Studies and University Studies)

Degrees offered

B. Metrics – Library Level
1. Size of library
   i. Number of Faculty and Staff (10/2013 data)
      • 24 tenured, tenure track and non-tenure track library faculty,
      • 11 administrative professionals (3 with rank & tenure),
      • 49 civil service staff

   ii. Operating Budget
      $9.2 million is the library operating budget, including personnel

   iii. Number of Volumes (10/2013 data)
      • 1,622,355 print volumes (located in open and closed stacks as well as a remote storage unit),
      • 81,448 electronic serials titles,
      • 1,927 print journal titles

   iv. Size of Digital Collections
      These numbers are estimated on the basis of digital masters.
      • Digital collections: ~2 TB (images and text, including books),
      • Archives: ~88.7 TB (negatives and books),
      • Cataloging and Acquisitions department:
      • Books and full-text from publishers: ~245 MB,
      • Newspapers and mixed formats: ~3 TB on external hard drives

2. Number of staff currently devoted to digital preservation: N/A
3. Library’s budget for digital preservation activities: N/A
4. Digital content management and preservation software, platforms, technologies, etc. that are currently in use
   • Archon,
   • CONTENTdm,
   • Digital Commons,
   • Internet Archive,
   • Faculty publication database
C. Defining Moment
   1. What happened that made you realize that DP was an urgent, unwieldy, and unaddressed problem at your institution
   Though digital preservation has been an important asset to consider for the ever-increasing volume of digital files, the IMLS NLG has triggered a certain level of interest within the library. Equally important have been the federal requirements for grant submission at the institutional level, and the recently enacted Illinois legislation (i.e., SB1900), which mandates that “each public university shall establish an Open Access to Research Task Force” by January 2014. Further, any digital transformation of unique analog collections held at Milner Library, which constitute an asset to teaching and research at Illinois State University, represents an investment in the long-term access of the institutional memory. We have increasingly become aware that there is no centralized plan to collect data from each department and that everything in Archives was stored on portable hard drives.

II. Self-assessment results (Please see <imls_digpres_selfassessment_isu.pdf> attached)

III. Policy Gap Analysis

A. “As Is” Model – where you are
   Though Milner Library has no digital preservation policy, it is beginning the work towards a digital preservation solution. The current practice includes three distinct, yet interconnected departments. These departments represent different facets of the organizational chart, each representing some aspect of Illinois State University history. They are the University Archives, Special Collections, and Digital Collections. The University Archives holds rich and valuable legacy materials from the community and the history of the university. While University Archives digitize their analogs internally and store them on local servers, the department outsources larger projects for digitization. One of those services is Internet Archive, which provides a storage component. Special collections hold rare materials including one of the top five Circus collections in the world, Lincolnalia and more. The Special Collections department coordinates their digitization initiatives in part, with Milner digital center. The Digital Collections department houses the digital center, where analogs are digitally transformed into Master copies and documented per NARA standards. Digital materials are stored on local servers and magnetic tapes and backed up daily. These collections and others are made accessible via the Digital Collections department website, available via a CONTENTdm Hosting instance. Lastly, the Digital Collections Librarian manages university scholarly and research materials by means of the institutional repository, ISU ReD.

B. “To Be” Model - where you want to be (different for each institution - in the spirit of established best practices)
   It is important for Milner Library to develop a digital preservation policy in order to carry out the mission of supporting teaching, learning, and research. This policy will provide guidance in determining which materials and institutional memory content to permanently preserve. The policy will address selection criteria, workflows and technical requirements for storage and 4
retrieval. Additionally, elements would include staffing and professional development to guarantee the success of a digital preservation initiative at Milner Library in agreement with Illinois State University stakeholders. Concerted efforts across departments and networking on campus will maximize resources and achieve digital preservation goals.

Several initiatives have recently highlighted the importance of digital preservation.
1) The IMLS digital POWRR grant has brought a minimal awareness on campus,
2) The institutional repository, ISU ReD, was launched in the spring of 2013,
3) The Data Management Plan requirement from the National Science Foundation (NSF), large National Institutes of Health (NIH) grants, and the digital humanities directorate of the National Endowment for the Humanities (NEH) have all raised the issue of digital preservation,
4) The Illinois Senate Bill 1900 has mandated that “each public university shall establish an Open Access to Research Task Force” by January 2014.
The assets that are collected via these initiatives represent a major step toward embracing the need to develop strategies and take action in digital preservation.

C. Gap – what specifically is keeping your institution from achieving the “To Be” (money, staffing, technical infrastructure, buy-in, etc.)
In the newly published 2013 report by Primary Research Group, Inc., Danielle Mericle, Director of the Digital Media Group at Cornell University states, “until we had a preservation repository in place, it didn’t make sense to embark heavily on digitization.” Milner Library considers it is important to identify the university collections for which long-term initiatives may be taken for granted if not intentional. Beagrie et al. (2008) recommend the careful examination of “dependencies and linkages between...preservation and collection development policies; the mission, preservation capacities and expertise of [the] organization; and scope your preservation policy and implementation plans accordingly” (p. 14).
In order to close the gap, Milner Library must literally build bridges between people and departments involved in the preservation work. These include the curators of collections, conservators, catalogers and those who would provide the digitization. In other words, closing the gap requires administrative support on many levels, buy-in from campus stakeholders, coordination with other offices on campus, priorities for our restricted resources, technical infrastructure and staffing. Fundamentally, all stakeholders must understand their common concerns and the importance of collaboration to achieve their goals.
In addition, though challenges to digital preservation may relate to economics and human capital, one fundamental misconception is to understand that preservation and access are different. Further, administrators need to “understand that digital preservation is not peripheral; it is a cultural change; an institutional activity” (Halbert, 2012 cited in Rinehart, Prud’homme, & Huot, 2014, p. 6), it encompasses a series of procedures and processes in sustaining long-term access and preventing further deterioration to the selected materials in support of the mission and goals of the library.

D. Digital Preservation Policy/Program Proposal
1. Mission/Scope
Protecting and ensuring accessibility of selected digital data to be collected and preserved in a secure and scalable environment, which practices adhere to technical standards and comply with legal retention requirements for extensibility and flexibility of the digital cultural heritage of the institution.
2. Etc. – Each institution can choose their own framework on which to base their proposed policy.
The proposed preservation policy would focus on the commitment to developing innovative services, programs, space, strategies, and systems that promote discovery, dialogue and learning to the constituencies. In this, it aligns with the mission of Milner Library “to create and sustain an intuitive and trusted information environment that enables learning and the advancement of knowledge in a culturally and technologically superior setting.” At this point in time, a digital preservation policy has not been linked to other internal library department strategies; however its alignment with University Archives and Digital Collections would solidify its integrity for a successful outreach to external entities on campus. Further, the policy would align with digitization strategies to reflect on the types of digital materials the library holds. Externally, the LEAPForward project on campus, a student information system, is in the process of data analysis of technology requirements. As data migration will be done by end of year 2014, collaborative analyses will be implemented. Other aspects of data management that have an impact at the campus level relate to federal requirements with NSF, NEH, and NIH grants submittals. These requirements play an important role for the long-term research prospects and commitment to keep original data to its highest quality with all inclusive metadata and maintain the user experience through the institutional memory. Also, the enacted Illinois open-access bill (SB 1900) mandates that “each public university shall establish an Open Access to Research Task Force” by January 2014. Therefore, it is essential to commit to “preserving the archival version of the resource in its highest quality in accordance to collection development policy retention decisions. The library will identify a set of priorities based on a series of principles pertaining to the commitment to long-term preservation of selected content, the relevant significance and uniqueness of materials, the guarantee that the data collection is reliable and the adequacy of the level of preservation of the deposited data the organization will engage in. Lastly, the level of integrity of the materials and legal obligations and varieties in file formats may influence the satisfactory level of preservation provided.

3. Digital Preservation Workflow – Specific to your institution and the types of digital objects you anticipate will be “in play”

i. Selection and acquisition
The selection policy will guide the selection of material, regardless of access platform and department. The policy will detail the relationship of the material to the goals of the university. By providing their content expertise, collection specialists will collect materials that are relevant to scholarship and history of the university. Materials accepted will range from donations to records’ transfers in a variety of formats relative to the particular computing environment to guarantee their accurate presentation in consultation with technical experts.

ii. Curation
The curation phase of the workflow focuses on the active management and appraisal of data over the entire life cycle. The policy will permit Milner Library to maintain and procure added value for those materials selected and collected for long-term preservation and access with consideration of standards and guidelines at the time digital assets are created. The Digital Center will guarantee that digitization standards are applied to promote the long term value of those assets in their original formats. The applicability of standards will extend to metadata creation, i.e. MODS, EAD, etc., including preservation metadata and appropriate information packages in accordance to the OAIS reference model to ensure interoperability across multiple systems.

iii. Storage and Access
Storage will be geographically dispersed and redundant with continual monitoring, not excluding scalability and affordability. Access of digital assets is to be reserved to designated personnel only. Access of digital assets will be handled by digital preservation personnel in accordance to whether materials are to be publicly available or rendered for research purposes only, i.e. non-public
materials. Theoretically, users could request specific folders of records or documents once selected from ARCHON finding aid or CONTENTdm for example. A separate site would be created for accessing selected copies to be downloaded directly without having to make use of the information packages created for access purposes only.

The structure of digital preservation practices will comply with access restrictions as defined in all relevant laws, regulations, licenses and deposit agreements.

E. Strategy – How will you direct people and organizations towards embracing and implementing the Policy

1. Communication & Education
   i. Administrators
      One-on-one meetings have proven useful to encourage buy-in and funding of resources.
   ii. Colleagues who are stakeholders
      All department heads (Archives, Special Collections, Preservation, and Digital Collections) will continue on-going conversations in and out the library. External offices to be included in outreach efforts may be the Office of Research and Sponsored Programs with respect to grants and research projects produced at Illinois State University. Further, the Graduate School is the main point of contact for Electronic Thesis and Dissertations and the long-term preservation of this intellectual capital at Illinois State University. Lastly, two employees have become certified as trainer in the DPOE (Digital Preservation Outreach & Education) trainer network.

   iii. Content providers
      Milner Library holds several collections, housed in Special Collections, University Archives. An array of collections also exists throughout the institution, which is being identified gradually with the support of librarian subject specialists.

   iv. General marketing/PR
      Generally speaking, the public relations and marketing office at Milner helps promote digital preservation information to stakeholders. At the same time, librarian subject specialists do help with direct contact to academic department, respective of their discipline.

2. Allocation of Resources
   i. People
      Temporary digital preservation non-tenure track position has been proposed to library administration.
   ii. Budget N/A

IV. Lessons Learned
Objectively, the writing of a digital preservation policy will be the product of an agreement between stakeholders in the library and the institution. A working group will be charged with the broad responsibility to ensure that the library’s policy and technical infrastructure can meet digital preservation priorities at the institutional level.

Observations from campus interviews indicate a largely non-uniform understanding of digital preservation and what it means to preserve materials for long term access, while many stakeholders are involved in day-to-day responsibilities in a learning environment, which also calls for immediate access to information. Further, observations are, that more outreach is needed to rally advocates for digital preservation to become part of the organizational culture.
References
    colleges/illinois-state-university-1692
I.A.3. Carnegie Classification,
    http://classifications.carnegiefoundation.org/lookup_listings/index.php
I.B. Milner Library, Illinois State University,
    http://library.illinoisstate.edu/libraryinformation/
I.C.1. SB1900: Public Act 098-0295, Enacted,
III.A. ISU ReD, http://ir.library.illinoisstate.edu/
III.B. ODU Libraries Digital Preservation Policy. Retrieved from
III.B.1. Preserving (Digital) Objects With Restricted Resources, digitalpowrr.niu.edu
III.C. Beagrie et al. (2008). Digital preservation policies study. Retrieved from
    http://www.jisc.ac.uk/media/documents/programmes/preservation/jiscpolicy_p1finalreport.pdf
    preservation challenges at the under-resourced institution. OCLC Systems & Services, 30(1). 28-42.
    ODU Libraries Digital Preservation Policy. (2012). Retrieved from
III.D.2.
    Mission of Milner Library. Retrieved from
    http://library.illinoisstate.edu/library-information/about/mission-vision.php
    The Ohio State University, Digital preservation policy framework
III.D.3.iii.
    Digital Preservation Framework, University of Minnesota Libraries.
    Preservation-Framework-Final.pdf
    Jisc Digital Preservation briefing paper
    http://www.jisc.ac.uk/publications/briefingpapers/2006/pub_digipreservationbp.aspx
I. Institution Bio & Back Story

Illinois Wesleyan University http://www.iwu.edu
1312 Park St.
Bloomington, IL 61701

Founded in 1850, Illinois Wesleyan has grown to become one of the nation's leading liberal arts institutions. The University enrolls just over 2,000 undergraduates from across the nation and around the globe and our graduates can be found in all 50 states and 52 countries.

Metrics: Institutional Level

Number of students, faculty and staff Facts 2013-14

Undergraduate Students: 2009
Faculty: 183
Staff: 263

Endowment and budget:

Endowment: $214 million (July 31, 2013)
Operating budget: $90 million

Carnegie classification:

Classification
Undergraduate Instructional Program:
Graduate Instructional Program:
Enrollment Profile:
Undergraduate Profile:
Size and Setting:
Basic

Category
A&S+Prof/NGC: Arts & sciences plus professions, no graduate coexistence
(not applicable)
ExU4: Exclusively undergraduate four-year
FT4/MS/LTI: Full-time four-year, more selective, lower transfer-in
S4/HR: Small four-year, highly residential
Bac/A&S: Baccalaureate Colleges--Arts & Sciences

Degree-granting Schools/Programs: College of Liberal Arts and Sciences, School of Art, School of Music, School of Nursing, School of Theatre Arts.

Degrees offered: BSN, BFA, BM, BME, BA, BS
Metrics: Library Level

Size of the library

Number of faculty and staff: 9 professional librarians; 10 staff

Operating budget: $1,110,387.00 (not including salary and benefits)

Number of volumes: over 300,000

Size of digital collections: approx. 451GB (as of Nov 2013). This is most of the library-held content and does not include campus photographers' collections or items ingested into our IR, Digital Commons, aside from scanned collections. There is some overlap due to the scanned content in the IR, but the total quantity in our IR as of January 2014 is 8GB.

Material not accounted for in either estimate is born-digital A/V material that goes into Vimeo or YouTube for ease of streaming access. These materials have not been appraised for long-term value as of this report. Streaming content that is accounted for (starting in AY 2011-12) in the off-line storage estimate is content that is broadcast through the campus subscription to a service called Stretch, which is used for real-time broadcasting of athletics and major campus events. This content is downloaded annually and added to the archives' off-line storage and so could be moved into a preservation system, but it is not being created in a recommended format. It is unlikely that creation practices for any A/V content will achieve "best practices" standards, so what we are currently receiving is a good representation of what we must consider in our planning.

At this time, I have determined that our born-digital A/V formats contain the content that is most at risk and in need of monitoring for format migration and file degradation.

Staff

Number currently working on digital preservation 2 FTE but not full time; one person is doing about 30% of time on this grant and one is at about .5%. If a formal preservation program is implemented, I hope we will be able to involve at least one other staff person (probably no more than 1% of time) to become the metadata wrangler and one student to accomplish ingest work.

Organizational role: University Archivist & Special Collections Librarian, Digital Projects & Reserves Coordinator.

Responsibilities beyond digital preservation: For archivist: liaison for two departments and an academic program (collection development, instruction and student research assistance needs) as well as faculty responsibilities for student advising, campus service and scholarship. For digital projects: reserves (e- and regular) and some IR development/maintenance work.

Budget

We currently have no funds allocated for digital preservation. If we chose to acquire a product through an annual subscription, discussion over funding would stay in the library unless wider responsibilities for campus material were involved.
Digital preservation technologies currently in use

None, although during this grant period the archivist evaluated and tested Archivematica as the most robust, promising pre-ingest tool available given our staffing capabilities. However, as of this writing (in March 2014) there are unresolved questions about some of the hurdles encountered with it and it is unclear if there would be on-campus ITS support for maintaining the Oracle VM Virtual Box needed to run the program. One preservation back-end system (DuraCloud) that offers limited—but multiple—user SIP and DIP access also seems like it would meet our needs but only for content that rises to a level of a full preservation system expense. We use two content management systems: Berkeley Electronic Press’s (bepress) Digital Commons (our institutional repository) and OCLC’s CONTENTdm (our instance of the latter is hosted by the Consortium of Academic and Research Libraries in Illinois—CARLI; see Section II for details on their uses).

Defining Moment

The defining moment for me, the university’s archivist, happened in about 2008 and involved the realization that our newly-subscribed to institutional repository was not a preservation system. It is an excellent backup choice and does manage some metadata for us, but there is no bit-level analysis on ingest or during storage and no means for detecting and replacing corrupted data. This realization came during a NEDCC workshop on digital preservation when the newly developing topic’s relationship to IRs was addressed. The more I learned about what is involved in digital preservation the less confident I became that my library could engage in all the activities required to live up to the standards. We did not have the training, staff or budget needed to implement a full preservation program. What I was able to do was begin educating my colleagues on the differences between storage and preservation and start raising awareness of format obsolescence issues elsewhere on campus. I also created and began maintaining an inventory of the digital objects I was responsible for.

II. Self-Assessment Results

Most of our work in digitizing collections has been accomplished through outsourcing. Vendors have supplied formats for both preservation and access as well as checksums for the digital objects they create. Our policy for digitization states we will retain the analog originals and make them available for use on site if needed. For some fragile paper material and outdated media types, it is our practice that the digital surrogates are the primary access copies.

Our content management systems are CONTENTdm and DigitalCommons@IWU. We have collection development, access and take-down policies for these materials. Our Scholarly Communications Librarian acquires non-exclusive licensing agreements for student-contributed content. Explicit permission is acquired for any non-IWU born-digital content that is accessioned by the archives.

We are hoping to implement a hosted version of ARCHON in 2014 to help us coordinate metadata for understanding and accessing our distributed collections. CARLI, the consortium that hosts the ARCHON instance we will be using, is discouraging members from viewing it as a repository for digital objects themselves.
Not everything that is digitized or born digital is accessible through online portals. Some materials are completely offline. Our collection development policies guide these distinctions. Objects in our CONTENTdm collections are completely open access; some parts of our IR are restricted to campus IP ranges but may be accessed from anywhere with valid credentials.

Master files of digitized materials are held on the deliverable media, if provided that way by vendors, and have also been copied to a 5-disk networked drive being used for off-line storage. One disk failed in 2014, five years after implementation, and the entire RAID was replaced. The drive is stored with other servers in a building that is on the same campus but not in the same place as the analog originals or disk media copies. There is no possibility of securing a geographically distributed back up location for the RAID at this time. Access copies that exist in both content management systems are in separate locations, and the IR materials have added assurance of protection through bepress’s geographically distributed server networks.

None of our current systems and policies includes opportunities for bit-level file degradation analysis or format migration. It is unlikely we will achieve this level of protection in the near future.

III. Policy Gap Analysis

A. “As Is” Model – where you are

For library-created/held collections we are in good shape to implement a digital preservation program. We know what we have and where the content is housed and have gained knowledge during this grant period about products and activities to accomplish all of the actions recommended in Level 2 of the NDSA Levels of Digital Preservation (http://www.digitalpreservation.gov/ndsaworking_groups/documents/Levels_v1.pdf).

Through the library’s IT budget, the archivist secured a five disk RAID configured external drive ca.2010; this is the off-line repository for digitized collections that the archivist selected for protection at any level. This portion of our collections represents the bulk of previously-digitized and born-digital material that have been inventoried and described. Our Scholarly Communications Librarian administers our institutional repository and policies exist to guide selection of faculty and student works as well as born-digital institutional records that it is necessary to make accessible online.

The early part of this grant program involved contact with faculty and staff about their digital object creation and preservation practices. Due to content creator’s necessarily narrow scope of interests, outreach and education on DP issues (e.g., good back up practices, consistent file naming, and use of widely adopted formats) will continue to be among the services offered to all in our community. At this time, this is the responsibility of the archivist but educational efforts are ongoing and it is believed that others across campus will become advocates.

B. “To Be” Model – where you want to be (different for each institution – in the spirit of established best practices)

For library-created/held collections more emphasis is needed on keeping up with what we create and the processes being used, and with tracking/documenting methods used to normalize or transfer content off of media. We also need to collect existing checksums, or create them for
objects that don’t have them, and collocate with the other object metadata. Then we need to make sure that these metadata files are duplicated and secured off site. This is a workflow and education issue for all library personnel creating/curating digital objects.

As a campus, we need to discuss what A/V records are being created and how, and we need to become consistent about all issues surrounding photographic images but especially file naming so that preservation of unedited masters will be useful. If we can’t trace a derivative back to a higher resolution version, it seems misleading to advocate for expending funds on preserving the low-resolution copies. We are not the only institution facing this issue, so perhaps someone working on digital forensics will resolve the problem in the future. In the interim, I believe we can achieve Level 3 on the NDSA scale and most of our important content will be adequately secured.

That leaves us with deciding what content we will secure with full preservation treatment. During the course of this grant period I came across a decision making tree developed at the University of Utah (http://lib.utah.edu/collections/digital/digital-preservation.php—currently being update by the authors); there is also the Digital Preservation Coalition’s tree (http://dpconline.org/advice/preservationhandbook/decision-tree/decision-tree-interactive-assessment) that will aid conversations on this part of our work.

C. Gap – what specifically is keeping your institution from achieving the “To Be” (money, staffing, technical infrastructure, buy-in, etc.)
Our real risk comes from the actions of others that I cannot control and feel I’ve made little headway in addressing. Our institutional history is in jeopardy due to default choices people don’t even realize they are making. We need to develop a sense of urgency that this discussion is important enough to take time away from everything else that people are doing. Our institution relies too heavily on posting born-digital content to our website and does not consistently transmit the same copy to the archives or use file naming conventions that would make it possible to automate harvesting with web-archiving tools.

If we can get people on board to discuss these issues, the next sticking point will be to get individuals to change their content creation practices and take ownership of DP as part of their regular work. For units that generate content that will need full DP treatment, cost-sharing out of diminished budgets will be the next hurdle. Until we reach that point, I don’t believe we can discuss who should be curating university-wide content or what the specific workflows will be.

Pre-ingest micro services will ultimately be required for library and university content, so a whole new set of staffing questions arise: training issues around metadata capture/creation, workflows for normalizing files and/or system-specific ingest needs. The staffing decision will also contribute to a decision about preservation system choices. The more we can pay, the more “turn-key” hosted systems can do for us. Budget increases are unlikely in order to accommodate this, so staff will have to share the burden. Current staff load within the library does not make it seem like campus-wide services are realistic to offer at this time. Quality control will be an issue if these tasks are dispersed across campus units.

D. Digital Preservation Policy/Program Proposal
1. Mission/Scope
The mission of the program is to preserve and sustain long-term accessibility to the unique
digital collections created, purchased or acquired by the library for which it has custodial or
curatorial responsibilities.

In consideration of our institution’s available resources, our preservation policy for digital
content aims to achieve the following objectives:
- preserving and providing continued access to digital material, both born digital and
digitized;
- ensuring that preserved digital materials are authentic;
- preventing damage and deterioration of the physical media by ensuring adequate
environmental control;
- creating digitized copies in formats that make future migration possible; and
- changing the format of digital materials to preserve their intellectual content if necessary.

We will make efforts to take the needs and desires of our community (including digital content
creators) into consideration when making preservation decisions; however, as long as the library
retains sole responsibility for caring for this content, we assert decision making authority for
- formats we will accept as is and which ones we will normalize, if applicable, to protect
the content;
- levels of preservation/protection needed (storage vs. long-term, bit-level management);
- access levels needed and who the gatekeepers are for off-line material; and
- preservation systems considerations (e.g., hosted or not, end-user accessible or not).

2. Digital Preservation Workflow – Specific to your institution and the types of digital objects
you anticipate will be “in play”
i. Selection: Any member of IWU’s community is encouraged to nominate digital
content for preservation or analog material for digitization, but actual selection decisions
will be based on collection development policies and in coordination with either a liaison
librarian or the University Archivist. Digitization actions will be accomplished based on
criteria that still need to be agreed on within the library.
ii. Acquisition: Once selection decisions are finalized, appropriate acquisition steps
include collecting creator-provided metadata (detailed on an e-records transfer form).
iii. Curation (i.e., managing files and metadata): Without the availability of a defined
preservation system, the archivist will continue to manage master files and metadata as
before: in an Excel file. We have not consistently created checksums for locally created
digital objects, and it seems this small step in the DP direction will be achievable.
iv. Archiving: Minimally, collecting objects and metadata using Duke Data Accessioner
can be used at this time. Collection into AIPs through the use of Archivematica will be
investigated further.
v. Storage: At a minimum, I’d like to establish a practice of creating one additional copy
and storing it out of the library with all content the library has curatorial responsibility
for. This will require ITS support and an added server.
vi. Retrieval: Also system dependent, but it is not the intention of the library to promise
that unmediated, online access will be possible for all content we curate. Our current
content management systems are adequate for our community’s needs today.
E. Strategy – How will you direct people and organizations towards embracing and implementing the Policy

1. Communication & Education to the following (at a minimum) is the responsibility of the University Archivist at this point, but all library personnel need to be educated about our DP capabilities. Any person who has contact with our community may be asked to describe our services and should have a basic understanding of the difference between storage and preservation of digital content and things we do to increase access.

   i. Administrators: The bulk of this work falls on Academic Affairs since both the library and IT are likely to be involved. If DP extends to the wider community, the Provost will need to be conversant on the issues so they can be discussed in the Cabinet. Even if Public Act 98-0295 (formerly IL SB1900) never touches private schools, it would be helpful for us to be aware of what competitors will be planning for.

   ii. Colleagues who are stakeholders: Library and ITS staff are the primary group as of now. Library colleagues in collecting areas need to agree on minimal metadata types to collect and a means of transfer/storage. Workflows need to be established for these practices. ITS has numerous administrative demands but has been receptive to issues surrounding DP and made positive moves in recent years (e.g., mirroring the campus website and expanding server backup practices off campus). We need more server space, though, specifically for the library’s RAID.

   iii. Content providers: Anyone on campus could be involved but the Communications Office (graphic designers, web content creators and all photographers) are of particular interest from an institutional history viewpoint. Faculty in our School of Music recently emerged as potential partners: their interest is in furthering access points for selected student, faculty and guest recitals, and they are receptive (in theory) to expanding their content capture practices to acceptable preservation formats.

   iv. General marketing/PR: Audience includes BOT, Cabinet, Chairs, individual faculty, staff and students; for staff members who reach our extended community, Advancement, Alumni Relations and Admissions will also be important to strengthen ties with. ITS staff reach many users on our campus and have been advocates for good data management practices for years; continued partnerships in securing selected portions of the digital objects being created means educating ITS staff on the kind of content that rises to the archives’ interests.

2. Allocation of Resources: It is assumed the library is solely responsible for all content it collects or creates, but it is not possible at this point to state exact staffing or budget decisions that will be appropriate if our responsibilities increase to other campus units.

   i. People – Not likely to expand beyond the archivist and one other person. Currently this is the Digital Projects Coordinator, but recent experiences with a Copy Cataloging staff member for ingesting archival material in the IR are encouraging. This work is added on to other duties and will require further negotiation and training if DP workflows are added. It seems likely that this person would replace tasks being handled, in part, by the Digital Projects Coordinator rather than augmenting them. Work study students should be trainable on some other workflows, too.
ii. Budget – Not likely to have a dedicated budget line unless a subscription service is deemed necessary. I am not recommending subscriptions at this time although DuraCloud seems most suited to our needs. DP selection decisions are my priority now.

IV. Lessons Learned
Much of what I’ve learned is incorporated into sections III A-C. There are two major weaknesses that prevent IWU’s full engagement in digital preservation activities at this time: 1) we lack a culture of records transfer to a central location for processing, and 2) we lack staff who are trained in metadata creation/capture.

Overall, there is a significant amount of work to do in getting campus buy-in that digital preservation is even an issue. Some inroads are being made, but the pool of proponents is limited at this time. People beyond the library and ITS need to know they have both the power and the responsibility securing our digital heritage or it may not happen.

My most rewarding activities during this grant period have been the conversations I’ve had with people about what they value now and what they think will be valuable in the future. Content creators have unlimited means of distributing their works and rarely think about the implications for the future. My work ahead involves training a cadre of people who understand the issues and can help spread the word. But at the same time I need to train people on good curation practices. My time remains my most restricted resource.

While our ability to work with commercial and non-profit tools and services during this grant period has been interesting and informative, no tool can do the hard work of making decisions about what matters to our institutions. Tool choices are what everyone seems to want to know about, and I place myself in that category two years ago. I have learned that behaviors are what the cultural heritage community truly needs to spend time on. Nevertheless, there are some preservation products that we can consider for collections we create and/or capture and that we can adequately describe.
NIU Case Study

February 2014

I. Institution Bio & Back Story

Metrics – Institutional Level

Number of students and faculty. NIU Fact Sheet (Last updated 11/1/2012)

Students: 21,869
  Undergraduate: 16,552
  Graduate: 4,984
  Law: 333

Faculty & Staff: 4,648
  Instructional faculty: 1,185
  Administrative and professional staff: 1,000
  Operating staff: 2,463

Endowment and budget.

Endowment: $57.1 Million? (Wikipedia)
Budget:

Carnegie classification. (Website here.)

- Undergraduate Instructional Program: Bal/HGC
- Graduate Instructional Program: CompDoc/NMedVet
- Enrollment Profile: HU
- Undergraduate Profile: FT4/S/HTl
- Size and Setting: L4/R
- Basic: RU/H

7 degree-granting colleges. Business, Education, Engineering and Engineering Technology, Health and Human Sciences, Law, Liberal Arts and Sciences, and Visual and Performing Arts


Metrics – Library Level

Size of library.

- Number of Faculty and Staff 32 professional librarians; 65 operating staff & admin (whole library)
- Operating Budget ~ $9 million per annum (varies)
- Number of Volumes (?) Over 2 million. ()
- Size of Digital Collections ~7-10 TB (not sure)

Staff.
• Number of staff currently devoted to digital preservation in some manner: ~6 FTE (includes students; no single individual is specifically dedicated to digital preservation) within the unit; .5 FTE outside administrative support

• Organizational role: The organizational roles of people whose duties include digital preservation include administrators, curators, cataloger/metadata librarians, graduate assistants, and student workers.

• Responsibilities beyond digital preservation: all staffers involved in digital preservation have responsibilities beyond digital preservation. Collection development, digitization, cataloging/metadata of non DP-related materials, programming & application development, curation, preservation of physical objects, committee work, faculty requirements, etc. Most staffers are devoting 20% or less of their time to digital preservation work at this point.

Budget.

Library’s budget for digital preservation activities $195,000 per annum (this # is likely derived from the human resource/ personnel costs in the Digital Initiatives Unit). We do not have a dedicated budget for digital preservation products, services, or staff at this time. The need for dedicated resources for digital preservation has been broached to library administration, to a relatively warm reception. It is our hope that resources will be allocated in the near future.

Digital preservation technologies currently in use.

• ARCHON (EAD finding aids)
• DSpace
• Fedora/Islandora
• Backups are available via campus ITS for librarywide staff servers and in-house backups for the Digital Initiatives Unit servers.
• Informal tools used departmentally include Gmail accounts and external hard drives. None of these have coordinated metadata.

Defining Moment

The realization that digital preservation was an urgent, unwieldy, and unaddressed problem came in 2008. The library submitted an NEH grant application for the digitization of dime novels, a major holding in NIU’s special collections department. The application was rejected on the basis of not having a long-term plan for preservation of the data generated through the grant.

II. Self Assessment Results

While we have some policies that address electronic materials in the library (including a disaster plan and some unit-level collection development policies), there has not been thorough coordination or standardization of same between individual units, nor have we globally addressed rights management issues. An active digital initiatives unit has been in place since circa 1996. It was originally run on a cost recovery/grant funding model, but there has not been a culture of documentation about unit practices.

The Digital Initiatives Unit was incorporated into the library’s permanent organizational infrastructure slowly over the past five years. However, there are still 7-10 TB of legacy digital
projects that need sorting in various stages of completion, in addition to other initiatives from other units such as Rare Books and Special Collections and the Regional History Center/University Archives that involve both digitization and born-digital materials. Numerous formats and file types are included in our digital collections, from text to audiovisual. Access is provided via the library’s website where required.

We have two content management systems: DSpace for our institutional repository, and Fedora for all other materials. Not all digital collections have yet been ingested into either repository. We have a significant backlog of scanned materials (mostly due to the need for metadata and a lack of robust bandwidth and server speeds).

In general, the library has been responsible for its own IT needs, but our new University President just hired a new Chief Information Officer, and we understand that some centralization of IT services will likely be on the horizon. Our budgets over the past 5 years have either remained flat or decreased in all areas.

III. Policy Gap Analysis
Where We Are

Scattered. We have numerous collections and practices and levels of documentation across numerous departments, as a result of legacy funding structures and a lack of coordination across units. The Digital Initiatives Unit began as a grant-funded, self-sustaining digitization unit, responsible only for their own staffing and infrastructure. Over a decade, the unit has been integrated into the library’s infrastructure, but many legacy projects and systems are still in the process of being converted or migrated to the library’s servers and systems. We also have Huskie Commons, an institutional repository pilot project, which has content and a management system (DSpace) but no long-term back-end storage and preservation solution.

In the meantime, both the Regional History Center/University Archives and Rare Books and Special Collections have been slowly accepting electronic manuscripts as part of their collection development activities, relying on local backup and informal cloud storage (i.e. a Gmail account) at best to keep files safe. Both RBSC and RHC have jointly implemented Archon, so we now have the possibility of managing and adding metadata to our files, but do not have a long-term solution for preservation-level storage and migration. Nor do we have a comprehensive web presence for our digital collections that are designed to be public. Campus wide, there is no plan for records management in terms of long-term storage of records relevant to University history to the best of our knowledge; any records management is handled only in terms of legal requirements.

Our challenge is to write unified policies and select tools that serve each constituent unit, but provide cohesion for the library as a whole.

Where We Want To Be

Ideally, NIU would develop/select one set of guidelines, workflows, and storage for all of our digital objects designated for long term retention (digitization unit born digital materials, RBSC e-manuscripts, RHC/UA records). We need a digital collection development policy that sets priorities for what gets digitized, who manages the materials, etc. We also need to further develop campus relationships that allow us to leverage what we know about the library’s needs
to help meet the needs of the entire campus. Since the start of this project, the Open Access to Research Articles Act (OARAA) was signed into law, requiring state-funded institutions to examine open access and come up with policies for possible implementation by January 1, 2015. The bill includes digital preservation as part of its remit, and POWRR team members at NIU have been strong advocates for making sure that open access policies also take long-term access into account.

The Gap

Staffing has been a problem. We have people with skills and knowledge, but there has been a lack of impetus to leverage those skills and knowledge in a speedy manner. There has also been a process of education for administrators and staff, getting from “we need to do this, really” to “resources go here.” This has partially been taken care of via the IMLS grant, and the impetus on our campus to solve the problem is growing with the passage of the Open Access to Research Articles Act (OARAA) in Illinois, which mandates investigating institutional repositories, open access, and long-term preservation of research produced on state-funded campuses. Nonetheless, we still have a ways to go to get from theoretical administrative buy-in to boots-on-the-ground implementation. We need to do a strategic reallocation of personnel resources as well as straight-up funding. Not all staff is willing to be reallocated or retrained, but job security means we have to wait for them to move on of their own accord.

Money has also been a problem. Purchasing turnkey software systems may be out of reach financially as an individual campus, and we have not had the staff to implement open access systems to the level that we want them. We are currently investigating possible consortial or collaborative solutions, but those negotiations take time, often years, while bit rot continues to happen. Library IT is separate from campus IT in our institution, and campus IT runs on a cost-recovery model. Building infrastructure and relationships between the two IT relationships is fraught because there is consistent push-and-pull over very limited resources. Our new University President just hired a new Chief Information Officer, and we understand that some centralization of IT services may be on the horizon in the near future. Our budgets over the past 5 years have either remained flat or decreased in all areas.

Digital Preservation Policy/Program Proposal

Mission/Scope

Northern Illinois University Libraries enhance the overall missions of the university by promoting excellence and engagement in teaching and learning, research and scholarship, creativity and artistry, and outreach and service. Our digital preservation program seeks to enhance innovation and the exploration of new ideas by implementing sustainable technologies that facilitate greater long-term access to digital scholarship and collections, whether digitized or born-digital. We document the life of the university and its production of historic, current, and future knowledge. We ensure the longevity and availability of these digital materials produced by our faculty, staff, and students, as well as those collected for our libraries, as is appropriate to achieve this goal.
Digital Preservation Workflow

Selection. Selection will be performed by the appropriate subject specialists and/or unit and department heads in accordance with established collection development policies for those collections and units (this includes Huskie Commons).

Acquisition. Initial acquisition of relevant born-digital files will be performed by the appropriate subject specialists and/or unit and department heads in accordance with established collection development policies for those collections and units. Additionally, scanning to create digital objects is performed by Dig Lab staff in accordance with their policies. Individual units and content creators are responsible for furnishing relevant files in appropriate formats for preservation to Digital Initiatives for long-term preservation.

Curation. (Managing files and metadata): File management will be completed in the selected system, governed by collections-based policies for availability, access, and metadata creation. Metadata will adhere to professional standards (i.e. MARC, MODS, EAD etc.) with a goal of single entry of metadata promulgated across multiple systems if necessary to promote metadata promiscuity. Processing of digital materials before submission for AIP conversion is the responsibility of the collection managers (e.g. SIP creation/Archon record creation/findings aids/MARC records where applicable). Preservation metadata will be system-generated (i.e. Archivematica/Curator’s workbench), based on policies set in coordination with collection managers and Digital Initiatives Staff.

Archiving. Digital Initiatives staff will be responsible for creating and managing Archival Information Packets in preparation for storage, in consultation with individual collection managers to ensure appropriate metadata assignment and access levels, providing packet-level metadata, and uploading packets to the selected storage solution.

Storage. Storage needs to be locally redundant and geographically redundant; accessible to relevant staff only; live networked storage (rather than static CDs/DVDS/Hard drives/external drives); scalable; affordable. We are still negotiating how library storage and campus storage will intersect; we expect cooperation between campus IT and the library to meet our needs.

Retrieval. Creation of DIP: will be dictated by collection managers in terms of formats, etc. Actual creation of DIP will be handled by Digital Initiatives Staff on the fly, by the repository system or by patron request. Creation of DIPs may vary based on whether the DIP is designed to be publicly available or furnished for research purposes only (i.e. not public). Theoretically, patrons would be able to request specific folders of records/documents after viewing an ARCHON finding aid, although the records/documents themselves would not be publicly attached to the finding aid in collections where materials are not in the public domain. A separate site created for serving access copies of public domain materials would allow patrons to download them directly, without having to make use of the DIP system.
Strategy

Communication & education.

i. Administrators: OARAA encourages resource allocation towards open access and long term storage. Need to educate to work from nodding to actual funding.

ii. Colleagues who are stakeholders: Office of Sponsored Projects. Faculty (need more outreach). Library subject specialists (more outreach). Graduate School (ETDs).

iii. Content providers: Regional History Center/University Archives. Rare Books and Special Collections. Huskie Commons. Digital Initiatives Unit. Southeast Asia Collection. Other campus content providers as appropriate (i.e. Honors, Research Rookies, etc.)


Allocation of resources.

i. People: Not-yet existing but necessary positions in bold: Scholarly Communications and Outreach Librarian (not yet hired). Curator of Digital Collections. Curator of Rare Books and Special Collections. Repository Developer(s)—we need another one. Dedicated (not shared)Server/System administrator(s). Curator of Manuscripts, University Archives. Metadata Librarian.

ii. Budget: We will need budget for: On-campus redundant servers. Dedicated hard-funded allocation for digital preservation activities. Regular budget for local servers/storage space and redundant campus backup. Budget for subscription fees for geographically disparate offsite digital storage, and possibly for a robust digital preservation management system.

iii. What Else?

IV. Lessons Learned

It's hard to write a DP policy without knowing which tools will be used to implement it, what financial resources and staff will be available when, and what kind of consortial work will be done, and by whom. We have a push-pull between: “tell us what you will need at a minimal level” and “we won’t know what we will need until we know how we are doing it, which is dictated by what resources we have.” At some point, we need to choose our systems and begin figuring out how to implement and fund them.

Seizing relevant opportunities to relate this project to initiatives administrators understand (like OARAA in IL, and FASTR and Data Management Plan requirements nationally) has been a great help. People understand responsible stewardship and management needs to happen, but they aren’t clear on how they can and should make choices that lend themselves to that responsible stewardship, i.e. self-curation of objects they create (e.g. we can’t save everything, no, really. YOU NEED TO CHOOSE. Before talking to us about storage needs.).

Based on our campus interviews, faculty are largely uninformed about this issue: they are so busy trying to survive day-to-day, but most of them come away from talking to us with a better understanding of what they need to do to ensure responsible stewardship of their digital legacies. We need to do a lot more talking to a lot more people.
WIU Case Study

April 2013

I. Institution Bio and Back Story

Metrics - Institutional Level

Number of students and faculty (Fall 2013)

Students: 11,707
Undergraduate: 9,873
Graduate: 1,834
Total Employees: 1,890
Faculty: 720
Administrative Staff: 317
Staff: 853

Endowment and Budget:

Endowment: $46 million
Total Budget: $241.6 million
Appropriated Budget: $52.76 million

Carnegie Classification:

- Undergraduate Instructional Program: Prof+A&S/HGC: Professions plus arts & sciences, high graduate coexistence
- Graduate Instructional Program: S-Doc/Ed: Single doctoral (education)
- Enrollment Profile: HU: High undergraduate
- Undergraduate Profile: FT4/S/HTI: Full-time four-year, selective, higher transfer-in
- Size and Setting: L4/R: Large four-year, primarily residential
- Basic Master's L: Master's Colleges and Universities (larger programs)

Four Degree Granting Colleges: Arts and Sciences, Education and Human Services, Business and Technology, Fine Arts and Communication


Metrics – Library Level

Size of library:

- Number of faculty and staff: 39 staff; 15 librarians; 4 graduate assistants; 2 administrators
- Total Operating budget: $5.01 million
- Number of volumes: Approximately 800,000
- Size of digital collections: Approximately 2 TB

Staff:
- Number of staff in Digitization Unit: 2. Additionally, approximately 40 hours of student work is allotted each week. Archives/Special Collections Unit faculty and staff also offer occasional assistance to the Digitization Unit. Nobody is assigned to specifically digital preservation as a responsibility.

- Organizational role: The organizational roles of people whose duties include digital preservation (broadly understood) include the Digitization Unit coordinator, one library specialist (metadata), the Director of Archives/Special Collections, and a handful of student workers.

- No single person is responsible for digital preservation. It is simply part of the workflow for the aforementioned employees. All staff members involved with digital preservation devote less than 20% or less of their time to digital preservation work.

Digital preservation technologies currently in use:

- ARCHON (EAD finding aids)
- ContentDM
- Library purchased cloud storage
- Campus wide IT backup (tape storage)
- External hard drives, personal storage space.

Defining Moment:

There was never one specific moment that I realized we had a problem with digital preservation. It was a gradual realization that as we continued to digitize materials in house (primarily photographs, maps, manuscripts) that we were creating our own problem. It was always in the back of my mind, but it was never serious enough for me to do anything about it. It was not until the opportunity to join this grant presented itself that I actually actively thought about our desperate lack of preparedness for short and long term digital preservation.

II. Self-Assessment Results

WIU Libraries has minimal policies regarding electronic materials in the library. We have a library-wide disaster plan, but digital materials are only given cursory attention. Our plan relies heavily upon campus IT, and we have no control over their disaster planning and readiness. We have been involved in digitization initiatives (mostly the scanning of photographs from Special Collections) since 2003. We started a stand-alone Digitization Unit in 2008, but it was given neither the budget nor adequate staffing to become a comprehensive digitization operation.

To date we have approximately 2 TB of digital materials in WIU Libraries. They are stored in an embarrassing mish-mash of locations, and we do not have a great grasp over the materials. Almost all of our digital materials are analog items we have digitized. We have very little born digital items in the Libraries at this time. (Though this promises to change in the future as we are asked to take over administrative control of non-current materials from our Visual Productions Center and campus PBS television station.)

We do not currently have a true digital repository system at our institution. We use ContentDM through our CARLI membership to present many of our digital collections, but we are aware that this is not a true preservation system. Other than that, our digital preservation network is
cobbled together though campus wide IT, internal library solutions, and external hard drives and cds. In short, these solutions are definitely not ideal.

Our university recently hired its first Chief Information Officer, and he has already shaken up the way IT is handled on campus, including centralizing many of the campus wide functions. It is unclear how this will play out long term, but the library is represented on the IT advisory board by a very talented individual who is able to articulate the library's needs effectively.

III. Policy Gap Analysis

Where We Are
In short: Wayward. We have not had enough coordination among our library units to centralize our processes and procedures as they relate to digital preservation. Digital initiatives began in the Archives and Special Collections Unit, and we relied upon our Systems Unit to assist with much of the back end work. This worked well enough to make the materials accessible to users, but preservation was never considered. When WIU Libraries added a dedicated Digitization Unit it was more because library administration knew enough to know that we needed some sort of centralized digitization center, but not really what it should look like and how it should function. The end result is a half-hearted attempt at a digitization program, and one which overemphasizes one aspect of digitization (the actual production of digital items) at the expense of another aspect (digital preservation.)

What WIU Libraries needs to do is to study the entire digital library cycle and allocate resources to each area more equitably and wisely.

Where We Want to Be
The greatest concern with using the plural 'we' in this heading is that I am not convinced that there honestly is a 'we' in this discussion. At this point it's more 'I', and I simply do not have the ability to get everything done that I feel is necessary. But I would say that my colleagues agree that we need much better policies and procedures for our digital initiatives. We need more comprehensive digital collection development policies, both for digitized analog materials and for born digital items.

Furthermore, we need to find a way to have our voice heard on a campus wide platform. This is even more so the case after records management was transferred to University Libraries in 2011. We do not have a mandate to act as I believe we should to fulfill our records management requirements, particularly as it relates to digital records. Finding a way to make our voices heard and to acquire the additional resources we need to follow the laws is a significant challenge, but it's a direction in which we must proceed.

We also need to be aware of digital preservation-themed legislative mandates coming down to the state universities. This is, thankfully, on the central administration's radar for the moment, and we need to keep the pressure on to make appropriate recommendations benefiting University Libraries as WIU's primary digital information warehouse.

The Gap
Not surprisingly, WIU's primary barriers to building and sustaining a proper digital preservation program are staff and resources. Collectively we probably have enough people on our current staff who have the expertise to handle the various aspects of digital preservation, but there has not been an institutional commitment to assign to those individuals the duties. With a 2013 change in administration at WIU Libraries we have a Dean who is aware of the problem and has
a genuine interest in exploring ways to solve the problem, but to date no faculty member has a
workload assignment specifically dedicated to our digital library. Two classified staff members
lead the program. Beginning in 2014 our Digitization Unit will report directly to a faculty
member, and this is likely to improve the staffing situation.

Financial resources are an entirely different kettle of fish. There is no way we have the resources
to buy a turnkey digital preservation solution. It’s possible it might work if we could negotiate a
consortial package and we were able to secure a commitment from our central administration. I
foresee no way in which we could absorb that expense from our current appropriation. We have
not investigated any open source solution due to the extensive programming and maintenance
requirements needed to get something up and running. Our greatest hope for establishing an
appropriate digital preservation program rests in our new dean being successful in lobbying
central administration for new resources to make it a reality. I, for one, am not hopeful this is
likely.

**Digital Preservation Policy/Program Proposal**

**Mission/Scope**

University Libraries identifies, collects, organizes, preserves and provides access to information
resources and services to support the University's commitment to academic excellence,
educational opportunity, personal growth and social responsibility. This same mission extends
into our digital initiatives. We are the designated depository of institutional knowledge, and
much of this information now comes digitally. WIU Libraries also has a special commitment to
preserving the history and culture of a sixteen county area in west central Illinois. We digitize and
make available to our region's communities important materials documenting daily life and
customs. We consider it our obligation to preserve these digitized materials for our regional
partners.

**Digital Preservation Workflow**

Selection. Selection will be performed by the appropriate subject specialists and/or unit
and department heads in accordance with established collection development policies for those
collections and units. In most cases this will be the Head of Special Collections, in consultation
with the Coordinator of the Digitization Unit.

Acquisition. Initial acquisition of relevant born-digital files will be performed by the
appropriate subject specialists and/or unit and department heads in accordance with
established collection development policies for those collections and units. Additionally,
scanning to create digital objects is performed by Digitization Unit staff in accordance with their
policies. Individual units and content creators are responsible for furnishing relevant files in
appropriate formats for preservation to the Digitization Unit for long-term preservation.

Curation. (Managing files and metadata): File management will be completed in the
selected system, governed by collections-based policies for availability, access, and metadata
creation. Metadata will adhere to professional standards (i.e. MARC, MODS, EAD etc.) with a
goal of single entry of metadata promulgated across multiple systems if necessary to promote
metadata promiscuity. Preservation metadata will be system-generated (i.e.
Archivematica/Curator’s workbench), based on policies set in coordination with collection
managers and Digitization Staff.
Archiving. Digital Unit staff will be responsible for the archiving of digital files, in consultation with individual collection managers to ensure appropriate metadata assignment and access levels, providing packet-level metadata, and uploading packets to the selected storage solution.

Storage. Storage needs to be locally redundant and geographically redundant; accessible to relevant staff only; live networked storage (rather than static CDs/DVDS/Hard drives/external drives); scalable; affordable. We will likely rely upon a mixture of library storage and campus storage; we expect cooperation between campus IT and the library to meet our needs.

Retrieval. This has not yet been discussed.

Strategy

Communication & education.

i. Administrators: Particularly central administration knows nothing about this issue.

ii. Colleagues who are stakeholders: Office of Sponsored Projects. Faculty. Library liaisons.

iii. Content providers: University Archives and Special Collections. Digitization Unit. Other campus content providers as appropriate.

Allocation of resources.

i. People: Director of Archives and Special Collections. Systems librarian. Digitization Unit Coordinator. A Scholarly Communications librarian (which we do not have) is desperately needed to coordinate all of this.

ii. Budget: We will need budget for: On-campus redundant servers. Dedicated hard-funded allocation for digital preservation activities. Regular budget for local servers/storage space and redundant campus backup. Budget for subscription fees for geographically disparate offsite digital storage, and possibly for a robust digital preservation management system.

IV. Lessons Learned

This entire process has served as a major learning opportunity for me. I learned how complicated digital preservation is, I learned how far off the radar it is for most people, and I learned how very much I am in the same boat as so many colleagues.

If and when we fully implement a digital preservation program at my institution it will require a level of user education I have never before encountered. From doing my campus survey to information conversations with colleagues it is crystal clear that very few people understand what digital preservation truly involves and how such a program looks.

I will be curious to see how digital preservation all pans out on our campus. There are still so many unknowns—will there be a statewide system, will we have to go it alone, will there be a sub-sect of institutions taking on this task? The entire process has been incredibly enlightening. I have looked at problems from entirely new perspectives and delved deeper into this singular issue than I ever thought I would.
From Theory to Action:
"Good Enough" Digital Preservation Solutions for Under-Resourced Cultural Heritage Institutions

A Digital POWRR White Paper for the
Institute of Museum and Library Services
August 2014

AUTHORS
Jaime Schumacher
Lynne M. Thomas
Drew VandeCreek

Stacey Erdman
Jeff Hancks
Aaisha Haykal
Meg Miner
Patrice-Andre Prud'homme
Danielle Spalenka
# Table of Contents

INTRODUCTION ........................................................................................................... 3

ABOUT US ....................................................................................................................... 4

MOVING BEYOND OUR LIMITATIONS .......................................................................... 4

OUR INVESTIGATIVE APPROACH .................................................................................. 6

OUR RESULTS .................................................................................................................. 8

  Archivematica .............................................................................................................. 8

  Curator's Workbench .................................................................................................... 9

  DuraCloud .................................................................................................................... 9

  MetaArchive ................................................................................................................ 10

  Preservica .................................................................................................................... 11

  Internet Archive ........................................................................................................ 12

MOVING TOWARDS THE FUTURE ................................................................................. 13

  Potential solution models ............................................................................................ 13

  Communities of practice and advocacy are central to our success .......................... 14

  Recommendations for the developer community ..................................................... 14

CONCLUSIONS .............................................................................................................. 14

ACKNOWLEDGEMENTS ................................................................................................. 16

APPENDIX A - Functionality Terms Defined ................................................................. 17

APPENDIX B - Recommendations for the Developer Community ............................. 19

APPENDIX C - Case Study Summaries ...................................................................... 20

  NORTHERN ILLINOIS UNIVERSITY ........................................................................ 20

  CHICAGO STATE UNIVERSITY .................................................................................. 21

  ILLINOIS STATE UNIVERSITY .................................................................................... 21

  ILLINOIS WESLEYAN UNIVERSITY ......................................................................... 22

  WESTERN ILLINOIS UNIVERSITY ............................................................................ 23
INTRODUCTION

Libraries, archives, museums, and other cultural heritage organizations collect, create, and steward a rapidly increasing volume of digital content. Both research conclusions and professionals' real-life experiences expose the inherent fragility of this content. Digital materials are at risk due to a number of factors:¹

- The storage media on which the materials physically reside can fail or become out-of-date and unusable (e.g. floppy, zip, and Jaz disks)
- The software programs with which they were created become obsolete, making older formats unreadable by newer versions (e.g. WordStar and Lotus 1-2-3)
- The basic bits that comprise digital content can become corrupt over time (e.g. bit rot)
- Mismanagement or loss of materials before being acquired by a repository
- De-contextualization due to lack of descriptive information or metadata

The cultural heritage and information science communities have developed guidelines, best practices, policies, procedures, and processes that can enable an organization to achieve high levels of digital preservation. However, protocols like TRAC certification are often challenging and complex. Purported digital preservation tools and services are developed, updated, and occasionally abandoned at alarming speed. Practitioners need to build awareness and secure resources to address the problem of digital fragility, but many who attempt to approach this challenge are left feeling overwhelmed and under-resourced.²

This is particularly true for professionals serving smaller institutions that are often faced with restricted resources. Small staff sizes, a lack of specialized expertise, dated technical infrastructures, and/or limited budgets create unique barriers for the professional tasked with stewarding digital content. When combined, these factors can create a seemingly insurmountable obstacle. Practitioners at smaller institutions often do not have time to stay abreast of the frequent developments in the field of digital preservation, may not have the expertise or technical infrastructure necessary to install and maintain complex software solutions, and frequently lack the funds to pay for complete, ready-to-use solutions that may exist. Faced with what seems to be an enormous undertaking, many peers serving at institutions with limited resources find themselves too overwhelmed to take the first steps. They are also in need of practical information with which to educate colleagues and administrators on the risks of digital content loss, advocate for necessary resources, and take initial technical steps to improve the preservation of their digital holdings. The results of this project’s investigation provide pragmatic digital preservation options, in the form of actionable and practical steps, for professionals grappling with these issues.

ABOUT US

The Digital POWRR team (Preserving digital Objects With Restricted Resources) is comprised of archivists, curators, librarians, and a digital humanist, from small and mid-sized Illinois institutions lacking significant financial resources due to a decade of state budget cuts. At each of our five disparate campuses, we have digital content that we know is vulnerable. Yet we have been unable to come up with programmatic and technical solutions to mitigate that risk. Team members have been calling attention to the risks for almost a decade and pushing towards a realistic solution while attempting to triage our materials. The POWRR institutions are members of Illinois’ statewide consortium of academic and research libraries (CARLI) which, while recognizing the challenges its members are facing, also lacks resources to address them.

Below is a glimpse into the makeup of our team, using most recent numbers available to us:

<table>
<thead>
<tr>
<th>Institution</th>
<th># of Students</th>
<th>Institutional Endowment</th>
<th>Library Budget</th>
<th># of Librarians/Archivists</th>
<th># of Volumes &amp; Digital Collections</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chicago State</td>
<td>5,700</td>
<td>$3.6M</td>
<td>$4M</td>
<td>15</td>
<td>475K &amp; 1.5TB</td>
</tr>
<tr>
<td>Illinois State</td>
<td>20,500</td>
<td>$69M</td>
<td>$9.17M</td>
<td>24</td>
<td>1.5M &amp; 2TB</td>
</tr>
<tr>
<td>Illinois Wesleyan*</td>
<td>2,000</td>
<td>$208M</td>
<td>$2.6M</td>
<td>9</td>
<td>300K &amp; 1.4TB</td>
</tr>
<tr>
<td>Northern Illinois</td>
<td>19,000</td>
<td>$72M</td>
<td>$9.52M</td>
<td>26</td>
<td>3M &amp; 7-10TB</td>
</tr>
<tr>
<td>Western Illinois</td>
<td>11,700</td>
<td>$46M</td>
<td>$5.01M</td>
<td>15</td>
<td>800K &amp; 2TB</td>
</tr>
</tbody>
</table>

* Illinois Wesleyan is the sole private institution represented in this project; the rest are state-supported.

Each institution produced a case study that provides more information on its background, the unique challenges it is facing, the composition of its digital collections, and the details of its technical infrastructure including content management systems and repository software currently in use (summaries found in Appendix C; full versions found on the wiki). The case studies also contain a self-assessment summary and review of current digital curation and preservation activities, if any. Team members highlighted the practices and policies they would like to see implemented. We performed a gap analysis by identifying the obstacles that have prevented our organizations from achieving the desired outcomes and created a plan of action for eliminating or maneuvering around those barriers. Acknowledging the obstacles we face, both as individual institutions and in common with one another, our institutions are better prepared to move forward with the development of successful digital preservation programs.

Common elements emerged from our gap analyses: a lack of available financial resources; limited or nonexistent dedicated staff time for digital preservation activities; and inadequate levels of appropriate technical expertise. Some of the case studies also mentioned a lack of institutional awareness of the fragility of digital content and a lack of cohesive policies and practices across departments as a contributing factor towards the absence of real progress.

MOVING BEYOND OUR LIMITATIONS

Early in our team’s investigative efforts, we discovered a fundamental misconception preventing many cultural heritage professionals (including some of us) from making meaningful progress towards the development of an effective program. We assumed that digital preservation is an either/or proposition; either an institution has implemented successful digital curation and preservation measures or it has not.
We came to realize that the opposite is true. Digital preservation is best thought of as an incremental, ongoing, and ever-shifting set of actions, reactions, workflows, and policies. An iterative approach means that practitioners don’t have to start by creating or selecting a comprehensive solution and making hard and fast technology choices to be used for the next 20 years. They can start by taking small steps to prioritize and triage digital collections, while working to build awareness and advocate for resources. It is appropriate to focus efforts on the activities we can perform in the next six to twenty-four months to steward our digital content, rather than wait a decade for a potential perfect solution. This may not be an intuitive approach for professionals in the cultural heritage sector accustomed to thinking in terms of decades and centuries, but to wait is to risk catastrophic content loss.

We base this approach in part on the National Digital Stewardship Alliance’s (NDSA) Levels of Digital Preservation.

“A work in progress by the NDSA, it is intended to be a relatively easy-to-use set of guidelines useful not only for those just beginning to think about preserving their digital assets, but also for institutions planning the next steps in enhancing their existing digital preservation systems and workflows. It allows institutions to assess the level of preservation achieved for specific materials in their custody, or their entire preservation infrastructure. It is not designed to assess the robustness of digital preservation programs as a whole since it does not cover such things as policies, staffing, or organizational support. The guidelines are organized into five functional areas that are at the heart of digital preservation systems: storage and geographic location, file fixity and data integrity, information security, metadata, and file formats.”

The NDSA Levels rubric approach makes it easy to determine where an institution’s technical infrastructures and current workflows stand within the framework. Most of the POWRR institutions have not yet reached Level 1 in some categories. This approach also leads to institutions recognizing discrete, incremental steps they can take to shift a single square to the right for just one of the functional areas, such as improving storage practices from stand-alone media (e.g. CDs) to networked or geographically distributed servers. Numerous resources discussing fundamental activities like these are freely available. Small steps move institutions closer to the end goal of stabilizing and preserving digital materials. Institutions should not wait to take Level 1 actions they are capable of today, while they determine how to move from level 2 to level 3 and beyond.

Most practitioners are capable of taking these fundamental actions with resources and skills currently at their disposal. Indeed, the notion that it is necessary to research all available tools and services exhaustively before taking any basic steps to secure digital content is yet another misconception that often prevents any progress from occurring. The number of options makes this impractical. The COPTR registry alone reveals hundreds of tools and services that address some aspect of digital preservation and/or curation.

---

6 Resources include the article, ‘Walk This Way: Detailed Steps for Transferring Born-Digital Content from Media You Can Read In-House’ by Julianna Barrera-Gomez and Ricky Erway, as well as the DP 101 page on the POWRR website.
Fortunately, practitioners can get started with simple, freely available triage tools while researching which of the more robust solutions will best suit their needs. The POWRR team recommends this approach as a bridge to help manage digital objects in a responsible way while we address our institutional cultures and begin laying foundations for suitable digital preservation solutions and life-long management activities. Indeed, many of the most robust and comprehensive tools and services assume that Level 1 triage activities (e.g. getting data off disparate storage media and minimal inventory and/or simple metadata creation) have already taken place.

OUR INVESTIGATIVE APPROACH

To address the overwhelming number of available tools at the time of funding, the POWRR team developed an evaluative rubric based on the intersection of the Digital Curation Centre’s digital curation lifecycle and the OAIS Reference Model. The final results of this initial evaluation of tools and services were mapped to a Tool Grid for ease of comparison. Figure 1 depicts the categories the POWRR team developed to map digital curation lifecycle activities to analog curatorial and preservation practices familiar to cultural heritage professionals.

Figure 1 – POWRR’s overview of the path to digital preservation

Our project team divided up a list of nearly seventy tools amongst ourselves, with each team member responsible for investigating five tools. Many of the tools and services listed on preservation registries only provide a small amount of information about the functionalities required to curate and preserve digital materials. These are often called microservices. Very few tools provide more comprehensive functionality, referred to in this report as macroservices. Microservice tools can help to prepare digital objects for ingestion and storage in a more comprehensive macroservice-based preservation system, and can be used easily in a workflow customized to address local needs and idiosyncrasies. Some tools and services require programming expertise. The costs of the tools/services vary from those that are freely-available via open-source communities, to those that are cost-prohibitive for smaller institutions. Different tools and services perform different functions in the digital curation lifecycle. To demonstrate the variety of functionality, Figure 2 provides a simplified overview of where a small number of tools and services fit on the spectrum.

---

7 To see an example of a collection being inventoried and accessioned using a free, simple ingest tool called Data Accessioner and a common spreadsheet application, visit the POWRR website.
10 Macroservices generally combine numerous microservices into a single user interface.
As we began investigating some of the more robust solutions alluded to above, the POWRR team consulted with our Board of Advisors, a group of 6 professionals known for their expertise and thought leadership in the field of digital curation and preservation. Our Advisors guided us in the selection of several tools and services for the team to test more deeply. The selections included front-end/processing tools, some services focused primarily on back-end storage and preservation, and some that performed functions across the spectrum. Only those technologies that were believed to be affordable and usable by our target audience were selected, based on what was available to us at the time and what we could procure short-term licenses for. They included both freely available, open-source solutions as well as vendor-based applications:

- Archivematica
- Curator's Workbench
- DuraCloud
- Internet Archive
- MetaArchive
- Preservica
- Archivematica + DuraCloud

Each of these tools and services were tested on at least 3 of the POWRR institutions' campuses, within a variety of technical infrastructures, and on differing computing platforms. Every institution used a subset of its own digital collections for testing purposes, employing a standardized rubric and evaluation form to report the results of their testing. The results can be found on the POWRR wiki.

One of the major challenges of this project was evaluating how these tools handled metadata. The tools we investigated often have an automated process for ingesting metadata that has already been created, and the ability to automatically extract or create metadata manually, to different degrees. We approached our testing with "from scratch" collections to explore the functionality of the extant metadata creation options in the tools as they existed, as we thought it best to begin with what the tools did as standalones before seeing how they interacted with other tools.

---

11 Liz Bishoff - Principal Partner at the Bishoff Group, LCC; Steve Bromage - Executive Director for the Maine Historical Society; Martin Halbert - Dean of Libraries and Associate Professor at the University of North Texas; Jerome McDonough - Associate Professor at the University of Illinois; Christopher Prom - Assistant University Archivist and Professor at the University of Illinois; and Amy Rudersdorf - Assistant Director for Content at the Digital Public Library of America

12 Unfortunately, technical challenges with installations and end-user testing of the software took up so much time that we were unable to move to the next phase in the testing period that explored how these tools interfaced with descriptive tools (like ARCHON or CONTENTdm) that might already be in use. An investigation into the interoperability of some of these tools and services would be beneficial and should be explored further.
OUR RESULTS

For each tool and service that underwent extensive testing by the POWRR team, we provide:

- Brief background information
- A summary of our experience using the product
- Pricing information (as of May 2014)
- A breakdown of functionality (see Appendix A for definitions of terms used in Figures 3-8)
- A brief review of our experience with its customer service and user support community

Archivematica

Archivematica is an open-source application that enables users to process digital objects from ingest to access, employing open standards in compliance with the OAIS\textsuperscript{13} functional model. We tested the freely available beta version 0.10. Archivematica is compatible with METS, PREMIS, Dublin Core and other best-practice metadata standards. It required the installation of a virtual machine to host the application locally that was then accessed by team members via their web browser. Overall, this product received positive reviews. However, if you do not know what a virtual machine is and do not have access to systems support or an IT department, then consider either utilizing Archivematica’s related consulting services for assistance or its newly released hosted version, or perhaps choose a different tool altogether.

![Figure 3 - Archivematica Functionality](image)

Project team members produced a total of six reviews of Archivematica. Of these, three noted its straightforward organization, relative ease of use and attractive features. One reviewer concluded, “This feels very much like a tool with which I could become comfortable and work routinely.” A fourth reviewer, working alone, found it promising at first impression, but joined a fifth colleague in ultimately finding it difficult to understand and operate. She wondered if a general knowledge of “computer systems, command lines, etc.” was necessary for its successful use. “For a typical librarian/archivist, this will seem overwhelming,” they concluded. Three reviewers found Archivematica able to work with larger files. Of these, one reported that the product failed to work well with several common file formats, a bug that has been subsequently fixed in release 1.0. A reviewer less pleased with Archivematica’s scalability reported that it took “a few minutes to transfer and ingest” a file of 98.2 MB.\textsuperscript{14} The fact that this tool required the installation of a virtual machine gave several testers pause, as they lacked the skills to do so themselves, and thus were reliant on IT staff. In some cases, testers had difficulty procuring necessary administrative rights in compliance with campus policies. One noted that the required virtual machine software crashed several times during testing.


\textsuperscript{14} We discovered these types of software do not run as quickly as typical desktop software.
Evaluators making use of customer support and online user groups reported good experiences, with more support often coming from Archivematica developers than product users. Still, there is a healthy amount of activity on this tool’s Google Group, indicating an active user community. Archivematica is a freely available, open source software solution. The lead developer of the tool, Artefactual, does offer some fee-based services, including installation, maintenance, training, etc. and all pricing is completely open and available on their website.

Curator’s Workbench
Curator’s Workbench is an open-source tool that was developed for use at the University of North Carolina at Chapel Hill as a collection preparation and workflow utility for digital archival materials. It runs on a desktop computer and helps to manage files before they are stored in an institutional repository or dark archive. It generates a METS file and MODS descriptive metadata elements that can be mapped to individual objects and folders.

In a total of five reviews, one found “the idea behind the metadata crosswalk is great,” but the process was complicated. A second found Curator’s Workbench unusable due to the lack of step-by-step instructions (although the developers do provide a YouTube video) and the fact that s/he did not possess the metadata expertise apparently required for the successful operation of Curator’s Workbench. Another noted that the product was “difficult to use” if one was unfamiliar with the METS and MODS standards. This reviewer also noted that the product’s crosswalk (between existing metadata and MODS) feature “does not work... if your data is not in the right sequence/order.” A fourth reviewer reported that they were “kind of disappointed. This tool assumes a lot more knowledge about metadata schemas than some users may have.” A fifth reviewer was generally dissatisfied with the product, reporting that s/he “had to keep going back to the instructions to use this program.... it didn’t seem to work as easily as I had hoped.” While likely an elegant solution to UNC’s needs, Curator’s Workbench did not seem to fit well with our project team’s workflows and capabilities.

While the lead developer of this tool was responsive to inquiries, he indicated that the product is not being actively developed for a new release due to lack of current resources. There is minimal activity within the user community’s Google Group. Curator’s Workbench is a freely available, open source software solution.

DuraCloud
DuraCloud’s preservation service brokers cloud-based storage\textsuperscript{15} from several providers and provides value-added services beyond storage. It is available from DuraSpace and, as of May 2014, offers storage with Amazon S3, Amazon Glacier, and the San Diego Supercomputing Center.

\textsuperscript{15} Cloud storage is a model in which data are stored on remote servers accessible via the Internet (aka the "Cloud")
In a total of four reviews, one tester found it to be simple and easy to use, and described it as "an immediate solution to the digital preservation problem for a small institution." A second reviewer found the product to be easy to use, with complete and clear documentation available. S/he went on to note, however, that DuraCloud’s “very minimal metadata... (properties and tags)” capacity produced “minimal functionality” for those seeking to add and manage metadata for digital objects. A third reviewer described DuraCloud as “an easy, intuitive product once all the set up was complete.” S/he noted the attractiveness of DuraCloud’s feature allowing the streaming of media objects stored within it, but found the product’s Sync tool to be problematic in that it was not flexible in its use and thus less-than-ideal for his/her workflow. Another reviewer found the product to be intuitive and appreciated the easy access to reports on the make-up and status of the collections stored in DuraCloud. S/he also found the bulk ingest tool to be very useful, stating: “I was able to start an upload of a rather large collection of videos and images before leaving the office for the day and it continued running overnight with no issues.”

The service was well-documented with instructions and video tutorials, and the virtual, real-time training was helpful. Customer service was responsive to inquiries. DuraCloud has open pricing available on its website. We have found their pricing to be very competitive, with options for storing 1TB of content with 2 cloud storage providers for under $2,000 per year (as of May 2014).

**MetaArchive**

MetaArchive is a community-owned, private LOCKSS network administered by the non-profit Educopia Institute, founded in 2004, providing dark archive storage. When joining the MetaArchive Cooperative, an institution provides server space for the hosting of other institutions’ materials, in return for offsite storage of their own materials. One of MetaArchive’s great strengths is that it conforms to all of the standards of a TRAC certified repository.

**Figure 6 – MetaArchive Functionality**

---

16 **LOCKSS** (Lots of Copies Keep Stuff Safe) technology provides for the geographical distribution of multiple copies of digital objects within a network of cooperating institutions and is based at Stanford University Libraries.

17 A dark archive does not grant public access and only preserves the information it contains.
Four of the five POWRR institutions tested MetaArchive’s collaborative membership model: one institution serves as the technical lead, hosting a staging server, to which the other partner institutions transfer their digital content, typically Archival Information Packages (AIPs), via file transfer protocol (FTP). To put it simply, an AIP is a folder or package of content that has been at least minimally processed and described, and is ready for long-term storage. The Cooperative provides a few potential methods to its members of content packaging for transfer and, ultimately, harvest. Once content is placed on the staging server, it is harvested into the MetaArchive network of participating institutions. This model requires a high level of technical expertise at the Lead Institution and the ability to host and maintain both a staging server and the LOCKSS server. For the purposes of this investigation, we ran a pilot of the staging and ingest process. We did not host a LOCKSS server, due to the incompatibility of the project timeline and the MetaArchive membership requirements. The support team at MetaArchive was helpful in walking us through the pilot and exposing technical issues incurred with our initial ingest attempts.

We tested MetaArchive using the Library of Congress’s Bagger tool for packaging content. Bagger’s sensitivity to file naming conventions presented a technical obstacle. One reviewer whose collection included nonstandard file naming conventions noted: “MetaArchive requires that a ‘find-bad-files.py’ script be run against folders [of content] before using Bagger... Resolving problems with file naming conventions proved to be a bigger problem for me than anything else in this project.” Bulk renaming utilities that may solve this problem are available, but were untested by the POWRR team at large.

Interaction with the MetaArchive Cooperative community revealed that one of the most useful aspects of membership is the mutual support it provides. This community of practice is particularly active in communicating with one another about their digital preservation endeavors.

![Collaborative Membership Model Diagram](image)

Indeed, some members indicated that being a part of the MetaArchive Cooperative was a primary reason for joining and that the actual preservation of materials was secondary.

The MetaArchive Cooperative openly provides their memberships fees and technology costs on their website: [http://www.metaarchive.org/costs](http://www.metaarchive.org/costs). Collaborative Members are bands of institutions that look and act like one unified member because they share a central server. This level of membership allows collaborating institutions to preserve their co-hosted content for a fraction of what it would cost to do so as individual members. ($377/member annually with 20 institutions sharing 2TB of space)

**Preservica**

Preservica is a vendor solution allowing users to process, ingest, provide access to, and store (using Amazon S3 and Glacier) digital materials using OAIS-compliant workflows. It is a hosted solution that requires a user to install one piece of software locally for uploading the content into Preservica.
Study team members provided favorable reviews of Preservica. Reviewers found the documentation to be clear and the vendor very responsive to email questions. Although one reviewer experienced initial difficulty installing the upload tool on his/her desktop, all ultimately found Preservica to be reliable, readily customizable and usable “out of the box.” One reviewer mentioned Preservica’s ability to gather content via web crawling as especially useful in his/her setting. S/he concluded that “this tool went above and beyond my expectations. The interface makes following the workflow very easy and manageable. Using this tool makes me more confident in my ability to do digital preservation, even if I still feel like a novice.”

Preservica offers robust training, both in-person (for a higher cost) and remotely (for less), and the customer service team is helpful and thorough. Over the course of the POWRR project timeline, Preservica has become more receptive to meeting the needs of smaller and mid-sized institutions of limited means. They have developed a particular pricing, delivery, and service model for institutions in this market segment, in part due to feedback from the POWRR project team and its advisors. Preservica now provides open pricing for the Cloud Edition, but requires a quote request for their Standard and Enterprise Editions. The Cloud Edition is available starting at $3,950/year (includes 100GB of storage) with some modest up-front training fees.

**Internet Archive**

Internet Archive is a free-use, online digital library that allows members of the public to create an account and then upload and download digital objects. As it provides individuals or organizations uploading digital materials with off-site storage, we believe that it can serve as a very basic digital preservation tool for very small organizations that hold the rights to their digital content (or the content is public domain) and have no additional funding or staffing options. The team felt that while the Internet Archive was a viable option for these particular types of cultural heritage organizations, it was not robust enough for organizations that have materials with access restrictions but must be preserved.¹⁸

¹⁸ POWRR did not test Archive-It!, a robust, fee-based archiving service provided by Internet Archive.
Several small, local cultural heritage organizations with which two POWRR institutions work provided reviews of Internet Archive’s usefulness as a digital preservation tool. Representatives of these organizations evaluating Internet Archive ranged from part-time volunteers possessing moderate Internet proficiency to a professionally-trained public historian. All reported that the Internet Archive’s availability for free use made it very attractive to their organization due to the limited availability of funds. Every reviewer found the Internet Archive’s main web page to be badly organized and confusing. Each of the three participating volunteer workers independently struggled to understand how they could take advantage of the Internet Archive’s digital preservation function and manage their content over time.

As a result of this experience, the POWRR team created a detailed tutorial with step-by-step instructions and screenshots to walk even the most novice user through the entire process. This can be found on the project website.

MOVING TOWARDS THE FUTURE

Potential solution models (a.k.a. One solution will not rule them all)

We have some general recommendations based upon institutional types and resources and the tools and services we examined in depth. Some POWRR institutions also considered the digital preservation problem in its own specific institutional context (see Appendix C), as well as exploring possible collaborative solutions.

For the smallest institutions, who do not expect to be able to add or redirect financial and/or personnel resources to the problem in the near future, we recommend beginning with the use of Data Accessioner for triage, and uploading public domain materials to Internet Archive for public access and long-term storage. This is especially useful for volunteer-run or minimally funded cultural heritage institutions.

For institutions that have some financial resources available for redirection, but do not expect to add personnel anytime soon, we recommend investigating Preservica. You may opt to invest in using a different back end storage service (like DuraCloud or MetaArchive) for your level of resources, rather than using Preservica’s Amazon-based services that are currently bundled with their processing tool. Institutions in this category may also find membership in MetaArchive particularly beneficial for its community of practice. Also consider investigating the use of Data Accessioner, paired with DuraCloud for off-site, trustworthy preservation.

For institutions that lack financial resources but have sufficient technical staff to take on additional work, we recommend investigating Archivematica, with an eye to using DuraCloud or Amazon Glacier as a back-end storage solution. Alternatively, joining an existing collaborative group in MetaArchive (or encouraging the creation of one with better-resourced institutions in your region) is a potentially fruitful model to consider.

For institutions that have staffing and financial resources available, Preservica’s services seem to be the best out-of-the box solution among the limited products we tested during the grant period. Preservica requires the lowest learning curve and minimal staff-side interventions. Institutions in this category may also find membership in MetaArchive particularly beneficial for its community of practice.

---

19 These organizations included: Joseph F. Glidden Homestead and Historical Center; Ellwood House Museum; Sycamore History Museum; and Sycamore Public Library’s Joiner Local History Room.
20 Potential collaborative models will be available in the fall of 2014 on the POWRR wiki.
Communities of practice and advocacy are central to our success
After three years of working together, the POWRR team has developed into a definitive community of practice, solving common problems in radically different environments. While we didn’t set out to do so, the lesson of the importance of collaboration and cooperation to move forward was not lost on us. This lesson was particularly crucial in driving us to approach other organizations and communities of practice, who, we were surprised to learn, were just as eager to speak with us.

We initially entered this project believing that we had to develop digital preservation programs for ourselves, as larger organizations had already solved the problem and had moved on to other projects. We could not have been more wrong on both counts. Not only were the larger organizations eager to communicate with us and help, we learned that many are still struggling with implementing basic digital curation and preservation infrastructures and policies. Smaller and medium-sized organizations needing digital preservation help are not as alone as we think we are. But we need to help each other while helping ourselves with basic triage activities.

Talking to our peers is not enough to move digital preservation forward. Advocacy across an institution is integral to digital preservation success. Lack of financial resources is often related to lack of institutional awareness: awareness leads to identifying priorities which allows for redirecting or acquiring needed resources. Even the smallest cultural heritage institutions can be complex organizations with numerous stakeholders (e.g. boards of trustees). Grassroots advocacy, through building a large, vocal network of people invested in solving the problem, is the most effective way to gain the attention of those in charge of allocating resources. In most cases, a new program takes shape and reaches implementation only when a critical mass of an institution’s community has come to agree that it is the appropriate way forward. Educating and influencing stakeholders, including faculty, staff, students, administrators, donors, etc. is therefore crucial to the successful development of a digital preservation program where none had previously existed.

Recommendations for the developer community
A few questions remained at the end of product testing. When developers and vendors were queried about these issues, the responses caused some members of the POWRR team to decide on an incremental approach rather than pursuing funding for a full preservation system. These observations represent unmet needs for institutions who will struggle to achieve “best practices” in a time when it is still difficult for our communities and funding authorities to accept the theory of digital preservation as being worthy of genuine concern. Please see Appendix B for recommendations aimed at product developers and their funding institutions. We believe the recommended efforts, and the accompanying rationale, will go a long way in helping organizations to secure funding for any level of preservation.

CONCLUSIONS

The fragility of digital materials presents a serious threat to libraries, museums, archives, and other cultural heritage organizations holding these types of collections. A rapidly growing number of professionals are aware of this risk and the need for action. They also know of the standards and best practices around digital curation and preservation, but many of them are struggling to decide how to move forward with addressing the problem effectively. Some institutions have created and implemented robust digital preservation programs; however, medium-sized and smaller organizations with fewer resources like those of the POWRR institutions are in an especially vulnerable position. Lacking the specialization of responsibility that comes with a larger staff and budget, practitioners often feel overwhelmed by the problem and unable to explore potential solutions. The significant time commitment required for merely
studying the problem and identifying workable solutions alone, among the many competing theories and recommendations, is out of reach for a majority of our peers. Many professionals in these organizations do not know how to take the next steps from awareness to action, and thus, the risk remains unmitigated.

We believe that small and medium-sized institutions with limited resources can in fact make progress towards the effective preservation of digital materials in the following ways:

1. Understand that digital preservation is an incremental process. Digital preservation is achieved through cumulative activities of increasing efficacy. It is time to embrace a “good enough” approach to digital preservation. Do not begin by aiming for designing and implementing a robust program with a detailed workflow for ingesting digital objects into a technical infrastructure fully in keeping with standards and best practices. Rather...

2. Focus on a set of discrete activities that can immediately yield higher levels of preservation, however modest. These can include inventorying your existing content, educating content creators, and designing an ingest workflow. Examples of the activities we chose to start with can be found on our project website. These activities will often give you better information with which to...

3. Examine your institution’s strengths and potential challenges to committing resources to its digital legacy. Understanding where you are, where you want to be, identifying roadblocks preventing you from getting there, and connecting with allies who can help you move further along the way is crucial to this process. Sole practitioners cannot do this alone. Planning and advocacy is just as important as directly managing materials. Finally...

4. When exploring more robust technical solutions, understand that selecting more than one tool or service may be preferable. Different levels of effectiveness at different price points can match up with different sets of materials your institution has identified as having different preservation priorities.

Digital POWRR team members each sought sensible workflows for our unique situations. We believe that collaborating within our communities and capturing sufficient levels of metadata using simple accessioning tools will position us to move beyond near-term security. We have investigated more robust solutions and believe that there are tools and services available within reach of small and medium-sized organizations with restricted resources. However, there is no one-size-fits-all answer for our institutions or even for different digital collections within our care. Nevertheless, we now have well-defined, manageable activities we can engage in using the resources we already have, while we explore reasonable implementation solutions for our institutions and advocate for the resources we will need in the future.
ACKNOWLEDGEMENTS

The ongoing success of the POWRR Project is due to the concerted efforts of professionals who have crossed institutional, state, and even national boundaries to work together towards common goals. Without the dedication of these individuals, the deliverables and outcomes of this project, including this White Paper, would not have been possible. The POWRR Leads would like to thank the following individuals:

**POWRR Team**
- Nathan Books
- Sarah Fraser
- Sharon Hu
- Martin Kong
- Amanda Miller
- Gayle Porter
- Amanda Rinehart
- Matthew Short
- Joe Thomas
- Katharine White

**Advisory Board**
- Liz Bishoff
- Steve Bromage
- Martin Halbert
- Jerry McDonough
- Chris Prom
- Amy Rudersdorf

**Supporting Leadership**
- Richard Darga
- Patrick Dawson
- Cindy Ditzler
- Michael Lorenzen
- TJ Lusher
- Karen Schmidt
- Sandra Toro
- Dane Ward

**And Others...**
- Dwayne Buttler
- Felix Chu
- Charles Larry
- Judy Michaelson
- Annie Oelschlager
- Seth Shaw
- Linda Wade
- Paul Wheatley

We are deeply grateful for all of your guidance and extraordinary efforts.
APPENDIX A
Functionality Terms Defined

**Copy:** Creates both a master and access copy automatically

**Fixity Check:** Ensures a file has not been changed using checksum or digital signature

**Virus Scan:** Checks for malicious programs and macros

**File Dedupe:** Checks for and ignores duplicate files

**Auto Unique ID:** Automatically generates a unique identifier for a single object

**Auto Metadata Creation:** Automatically creates information about the object based on data that is embedded within the object

**Auto Metadata Harvest:** Automatically harvests information about the object that is located in an external repository, catalog, etc.

**Manual Metadata:** Allows for manual addition of information, including descriptive metadata, at record or batch level

**Rights Management:** Allows for establishment of an object's access, use, and alteration rights

**Package Metadata:** Packages metadata in a standards-based format

**Auto SIP Creation:** Automatically generates an OAIS compliant Submission Information Packet

**Public Interface:** Provides end-user access and functionality to objects

**Auto DIP Creation:** Automatically generates an OAIS compliant Dissemination Information Packet

**Auto AIP Creation:** Automatically generates an OAIS compliant Archival Information Packet

**Reliable, Long-term Bit Preservation:** Verifiable bit-level preservation for a period of greater than seven years

**Redundancy:** Keeps more than one copy of the object

**Geographically Dispersed Data Storage Model:** Keeps more than one copy of the object in more than one geographical region
Exit Strategy: Has a plan in place for the mass extraction of all preserved objects

Migration: Converting the object from one format to another format considered to be of greater stability

Monitoring: Automatically checks for object corruption

Auto Recovery: Automatically overwrites the corrupted object with an uncorrupted redundant copy
APPENDIX B
Recommendations for the Developer Community

1. All POWRR partners and likely all of our peers have invested over a decade in creating digital content from analog holdings. Those who did this according to best practices have also collected and stored checksums for these objects. Therefore, we need a way to test checksums generated in earlier digitization projects during accessioning workflows for preservation storage environments. The time lag between these processes indicates a theoretical possibility that some digital content has already degraded. We need to know if what we’re ingesting now is of good enough quality to preserve.

Rationale: Collections managers need to assure funders that their content is safe and worth the expense of long term storage. Vendors who promise file fixity checks only on ingest may be blamed if content retrieved at a later date is not usable. The chain-of-custody from digitization to long term preservation needs to satisfy each party that the quality is acceptable at the time of transfer.

2. It seems a reasonable thing to ask companies that are analyzing the health of our digital objects how often “self-healing” has occurred to date. Over a decade has passed since the preservation industry began. Only one company consulted during our testing period was able to state how many files have suffered from “bit rot,” that is, file degradation at the binary level. That number was zero for a company that’s been operating preservation storage for only three years. Companies with more file fixity experience could not answer the question. Selling the concept of bit rot is difficult at a theoretical level let alone when arguing for a portion of diminished budget lines.

Rationale: Collections managers need to prioritize the levels of care their content receives, and data about high risk versus low risk formats or the way that specific formats age would help establish priorities. If trends per format are not detectable that way, perhaps estimating how many digital objects per 100 (or 1000, etc.) have developed bit rot is possible. Vendors would value this level of analysis because collection managers would have enough factual information when making a case for budget expenditures.

3. Some of our digital content is held in institutional repositories. Not all repositories make it possible to differentiate between content that is new and content that is old. For those of us who cannot accommodate LOCKSS-based systems, which were created on the principle of crawling network-accessible data directly and comparing with previously ingested data, another form of analysis is needed. When content is transferred in bulk, either into a system like the consortial LOCKSS network we tested or into a sole-institution preservation storage environment, the ability to analyze content in bulk and extract only newly accessioned material would be a useful innovation.

Rationale: Collections managers cannot afford storage costs that include duplicate copies of previously ingested material. These people also cannot afford the time it would take to manually differentiate the content they receive in bulk to previously transferred material. Development of de-duplication processes on ingest or in pre-ingest accessioning workflows is needed to assure managers that they are only paying to store unique digital objects.
APPENDIX C
Case Study Summaries*

*Entire Case Studies located at http://powrr-wiki.lib.niu.edu/index.php/Case_Studies

NORTHERN ILLINOIS UNIVERSITY
Northern Illinois University (NIU) is an institution of approximately 19,000 students, with an endowment of $72 million. Its library employs thirty-two professional librarians, administers a budget of some $9.5 million per year, and contains over three million volumes. Although a number of librarians and other professionals devote portions of their time to digital preservation activities, representing a budget of approximately $195,000, no single staff member's activities are devoted specifically, or even principally, to digital preservation. Likewise, the library has not developed a digital preservation policy. NIU Libraries possess some 7-10 terabytes of unique digital materials, have stored and managed them in-house, and employ several software applications that enhance their preservation in some way, including ARCHON, DSpace, and a Fedora Commons repository that is not yet fully populated. None have coordinated metadata.

NIU project representatives realized importance of digital preservation upon the 2008 rejection of a grant proposal on the grounds that the institution lacked a viable means of preserving the digital materials to be produced. Their recent review described existing preservation practices as "scattered" and uncoordinated. Looking forward, digital preservation advocates at NIU hope to develop a policy comprised of a single set of guidelines and workflows, and provide a single storage location, for all digital objects designated for long-term retention. The gap between the present state of affairs and the desired situation exists in part due to a lack of personnel devoting meaningful amounts of time to digital preservation activities alone. This shortage largely stems from a situation in which a steadily decreasing amount of available funds are devoted to the provision of several traditional library functions and services at levels prohibiting the transfer of resources to the development of the technical expertise and infrastructure necessary to support effective digital preservation activities. A lack of available funds has also precluded any potential purchase of digital preservation utilities available from vendors.

Lessons learned include an awareness that despite a general lack of a comprehensive review of existing digital materials and available solutions, it is necessary to act swiftly on this matter in light of federal and state mandates. Lessons learned also include a realization that faculty and administrators are largely ignorant of the threat of digital data loss. Thus a strategy for establishing effective digital preservation measures on campus consists of educating these stakeholders about the threat of digital object loss and persuading them to devote available resources to mitigate it.

Potential Implementation Model: If project team members had to make a recommendation to our leadership team today, based upon current resources, we would look at a combination of services that work with our current infrastructure. DuraCloud would be our selection for back-end preservation, tied to a Preservica instance used to manage high-impact, high-priority born-digital materials from Rare Books and Special Collections, the University Archives, and the Regional History Center. The Digitization Lab would leverage their current instance of Fedora/Islandora that is already live to create archival packets of digitized materials for deposit in the same DuraCloud instance.

While we think this is a viable stopgap/short-term solution, we would prefer to investigate a statewide collaborative storage solution that interfaces with multiple front-end services for economies of scale. A statewide collaborative solution for storage would allow us to be part of a community of practice attuned
to our local needs and budgetary restrictions for long-term planning, which is, in our opinion, possibly even more important than which tools are selected for use, based on our findings from investigating MetaArchive.

**CHICAGO STATE UNIVERSITY**

Chicago State University (CSU) enrolls approximately 5700 students. The university's endowment is approximately $3 million. Its library employs fifteen librarians and professional staff members, administers an operating budget of approximately four million ($4M), and contains over 475,000 volumes. It has no digital preservation policy and its budget for professionals' digital preservation activities is made up of small portions of the time of four individuals already devoted to other responsibilities. CSU library's digital collections and archives presently hold approximately 1.5 terabytes of digital materials. Technologies currently in use that contribute to the preservation of digital objects in some way include Eloquent Archives, which includes information about archival collections; CONTENTdm, a content management system administered by the state library consortium; electronic theses and dissertations administered by ProQuest; and Internet Archive.

The defining moment when several library staff members recognized the importance of digital preservation activities occurred when they realized that grant activities digitizing library collections included no provision for storage or preservation. A review of present activities finds that CSU Libraries has not implemented any digital preservation measures. In light of the fact that the Libraries are moving forward with the digitization of large amounts of video and audio materials, CSU Libraries' lack of digital preservation capacity is especially important. Looking forward, digital preservation advocates at CSU seek to develop a digital preservation policy (presently in draft form) describing which materials will and will not be preserved, as well as guidelines and workflows defining how preservation measures are to take place. They also report a need to identify the existence and location of appropriate storage media for the resulting collections.

Closing the gap between the existing situation and the desired state of affairs will require financial resources devoted to staffing and education. While present staff members taking part in this study are well-informed about digital preservation issues, they lack the time necessary to transfer their knowledge into new digital preservation activities. Lessons learned include a new awareness of the high degree to which faculty members expose their digital materials to risk of loss. They also include an awareness of the progress that CSU Libraries have made toward the goal of enhanced digital object preservation, made up largely of staff members' increased awareness of the problem and potential solutions to it. As CSU Libraries work with a limited budget, project contributors at this institution urge administrators to repurpose and retrain existing staff members to address the digital preservation issue. The implementation of a new digital preservation policy in this way will require the education and persuasion of faculty members and administrators responsible for the allocation of staff resources.

**ILLINOIS STATE UNIVERSITY**

Illinois State University (ISU) is an institution of some 20,500 students, with an endowment of over $69 million. Its library employs twenty four professional librarians, administers a budget of over $9 million, and contains over 1.5 million volumes. The library has no digital preservation policy, nor a budget devoted to digital preservation activities, and no faculty or staff members presently devote any part of their efforts to them. ISU Libraries possess roughly two terabytes of materials identified as digital collections, three terabytes of other digital materials described as newspapers and mixed formats, and 89 terabytes of archival materials. Digital content management and preservation software presently in use include
ARCHON, CONTENTdm, Digital Commons, Internet Archive, and a faculty publications database. Individuals now acting as digital preservation advocates within ISU Libraries report that the invitation to take part in the present study first brought local attention to the subject.

The announcement of federal mandates requiring the public, online availability and preservation of research data, as well as a new Illinois law that directs state institutions to devise a plan for the provision of online public access to, and preservation of, research data has increased local interest in the digital preservation issue. A review of existing digital preservation activities at ISU Libraries describes the storage of digital materials on local servers backed up daily via magnetic tape.

Looking forward, digital preservation advocates at ISU hope to develop a policy guiding the selection of materials eligible for preservation measures and describing the workflows and technical capacities required to implement these measures. Lessons learned include a new awareness of university faculty members' varying understandings of the threat of digital object loss and their preservation. The development of a digital preservation policy will depend upon stakeholders in the library and the larger university reaching a common understanding of the problem achieved through outreach activities and dialogue. Implementation of a policy closing the gap between the present situation and the desired state of affairs largely hinges on the devoting of staff members' time to digital preservation activities; the provision of ongoing professional development opportunities for them; and the coordination of their activities.

ILLINOIS WESLEYAN UNIVERSITY
Illinois Wesleyan University (IWU) is a private institution enrolling over 2000 students, with an endowment of $214 million. Its library employs nine professional librarians, contains over 300,000 volumes and administers an operating budget of over $1.1 million. The Library possesses digital collections of over 450 GB. IWU Library assigns no budget funds to digital preservation activities, but one member of the Library faculty devotes 30% of her/his time to this study, and another devotes .5% to the development and maintenance of an institutional repository. The Library makes use of two technologies that contribute in some way to the management of digital objects: bepress' institutional repository Digital Commons and CONTENTdm provided by the state library consortium.

The University Archivist describes the moment at which the digital preservation problem became apparent as occurring when she realized that the above repository was not an effective preservation system, in that it offered no means by which staff members might discover corrupted or degraded data and replace it with an intact version of the file. At present Library faculty members have knowledge of what digital materials their collections contain and where they are stored, but the Library has not developed or implemented a digital preservation policy. Looking to the future, digital preservation advocates at IWU hope to establish procedures and processes by which they might manage digital collections, including normalized versions of files. They also seek to collect existing checksums or create them for objects lacking them, and attach these checksums to existing metadata for digital objects. In addition, they wish to establish criteria for the selection of materials appropriate for preservation measures, and devise a means by which they might store duplicate copies of all materials in an off-site location.

The gap between IWU's current situation and its desired outcome consists in part of a lack of awareness of digital preservation issues on the part of faculty members and administrators at the institution. Shrinking budgets also present a considerable obstacle to the retention of new staff members who might make digital preservation activities their principal focus and/or the acquisition of turnkey solutions requiring less staff labor and expertise. Lessons learned include an awareness of the fact that IWU lacks an institutional culture that values the transfer of records to units and personnel able to secure their enhanced preservation. In addition, the Library generally lacks staff members trained in metadata
creation and capture. In order to close this gap, digital preservation advocates at IWU have drafted a proposed policy asserting the Library’s authority to make determinations as to which formats and file types will be accepted for preservation, the levels of preservation measures appropriate for different types of materials, what access to preserved/off-line materials members of the university community may enjoy, and which preservation technologies are to be adopted. In addition, the proposed policy outlines a workflow for the preservation of digital materials and standards for their preservation. In order to achieve this goal, digital preservation advocates plan to engage in education and lobbying activities with administrators and other stakeholders, including content providers. Current budgetary constraints suggest that no new staff members devoted primarily to digital preservation activities are forthcoming. Thus current personnel will be asked to add digital preservation activities to their present work load.

Potential Implementation Model: If Illinois Wesleyan had to make a choice today, based upon current resources, Data Accessioner would be used to process digital objects sent on removable media to the archives. Steps needed to implement Archivematica locally would be explored further, and DuraCloud would be used to store content selected for long-term retention.

WESTERN ILLINOIS UNIVERSITY
Western Illinois University enrolled over 11,700 students in the fall semester of 2013. Its endowment is $46 million. Western Illinois University Libraries employ fifteen librarians and twenty-four staff members. The libraries contain approximately 800,000 volumes, have a digital materials collection of approximately 2 TB, and an operating budget of $5.01 million. At present Western Illinois University Libraries do not have a digital preservation policy in place. Two members of the Libraries’ Digitization Unit staff devote less than twenty percent of their working effort to digital preservation activities, and Archives/Special Collections Unit faculty and staff members add occasional assistance. Technologies contributing to the preservation of digital objects that are presently in use include ARCHON and CONTENTdm. The Libraries also secure digital objects through the use of purchased cloud storage capacity, campus-wide IT backup to tape drive, external hard drives, and individuals’ personal storage space. Although WIU’s POWRR Project Coordinator realized that that the Libraries were digitizing materials without thinking about their long-term availability, he did not come to grips with the implications of this situation until he began project work. Western Illinois University Libraries report that no coordinated digital preservation activities are currently in place.

Looking forward, there appears to be a need for a coordinated set of digital preservation policies and procedures, as well as the wherewithal to inform the broader campus community of the risk of digital object loss, in the Libraries. A recent turnover in library administration has brought a new awareness of the digital preservation issue, but a lack of staff and financial resources currently stand as a considerable obstacle to the Libraries’ attempt to build a digital preservation program. Lessons learned include a new recognition of both the risk of digital object loss and university faculty and staff members’ general lack of awareness of it. The development of a new digital preservation program at Western Illinois University will thus depend upon the success of a large-scale effort to educate faculty and staff members, as well as administrators, about the dangers at hand, and motivate them to take action.
From Theory to Action:  
A Pragmatic Approach to Digital Preservation Strategies and Tools  
A Digital POWRR Workshop - 2014

Workshop Expected Outcomes

- You will understand that different digital preservation tools/services can perform different functions within the digital curation lifecycle and be able to explain how these tools/services can be used within your institution's workflow.

- You will practice some of the initial pre-ingest activities* required to accession a digital collection and gain the skills necessary to repeat this process at your institution.

- You will gain hands on experience with a basic digital preservation tool and understand how it can be used within your institution's workflow.

- You will take away resources that help align communication and advocacy, as well as guide policymaking and tool selection/implementation.

- You will create a 3-3-3 Action Plan to implement in the following 3 months that will move you closer to your digital preservation goals.

* as described in the OCLC report "Walk this Way"

Personal Equipment Requirements

In order to participate fully please bring a laptop (MAC or PC). We will be running programs and working on sample digital collections located on USB drives (provided) and Tablets will not allow for this. Also, one of the exercise tools needs Java in order to run. Most laptops already have this application, but you can acquire a free version at: http://java.com/en/download/index.jsp. To check if your laptop has Java already installed, run the test at: http://www.java.com/en/download/installed.jsp.
Pre-Reading Requirements

1) Our first activity together will involve the NDSA Levels of Preservation. Please review this document and try to place your institution's current position on it.
http://www.digitalpreservation.gov/ndsa/working_groups/documents/l.evels_v1.pdf

2) Articles
   You've Got to Walk Before You Can Run (a high level view of the minimum steps needed to make digital preservation activities operational)
   http://www.oclc.org/content/dam/research/publications/library/2012/2012-06.pdf?urlm=168601
   Walk This Way (a detailed process for implementing DP activities. Please read the introduction for each section only—no tool installation or pre-workshop practice needed!)

AFTER the Workshop

The IMLS requires us to report on the impact of our work. In order for us to determine if we've been able to help you begin digital preservation activities at your institution we will be conducting follow up surveys (and possibly phone calls) three months after our workshop. We will have a series of questions for you to answer based on the workshop takeaways.

OPTIONAL Resources

There are many technical articles on digital preservation. We do not assume you have read any of them. If you feel you need to refresh your knowledge on this topic, two places you can go are:

Library of Congress DPOE
http://digitalpreservation.gov/education/curriculum.html

The POWRR website contains selected articles and sources we recommend for self-education.
http://digitalpowrr.niu.edu/digital-preservation-101/

PLEASE NOTE: Our intention in highlighting specific digital preservation tools and services during this workshop is to share our discoveries based on our specific institutional types. The opinions of the presenters may not be suited to every institution's setting and do not represent endorsement by the Institute of Museum and Library Services.
# Functionality Terms Defined

<table>
<thead>
<tr>
<th>Ingest</th>
<th>Processing</th>
<th>Access</th>
<th>Storage</th>
<th>Maintenance</th>
</tr>
</thead>
</table>

**Copy:** Creates both a master and access copy automatically

**Fixity Check:** Ensures a file has not been changed using checksum or digital signature

**Virus Scan:** Checks for malicious programs and macros

**File Dedupe:** Checks for and ignores duplicate files

**Auto Unique ID:** Automatically generates a unique identifier for a single object

**Auto Metadata Creation:** Automatically creates information about the object based on data that is embedded within the object

**Auto Metadata Harvest:** Automatically harvests information about the object that is located in an external repository, catalog, etc.

**Manual Metadata:** Allows for manual addition of information, including descriptive metadata, at record or batch level

**Rights Management:** Allows for establishment of an object's access, use, and alteration rights

**Package Metadata:** Packages metadata in a standards-based format

**Auto SIP Creation:** Automatically generates an OAIS compliant Submission Information Packet

**Public Interface:** Provides end-user access and functionality to objects

**Auto DIP Creation:** Automatically generates an OAIS compliant Dissemination Information Packet

**Auto AIP Creation:** Automatically generates an OAIS compliant Archival Information Packet

**Reliable, Long-term Bit Preservation:** Verifiable bit-level preservation for a period of greater than seven years

**Redundancy:** Keeps more than one copy of the object

**Geographically Dispersed Data Storage Model:** Keeps more than one copy of the object in more than one geographical region

**Exit Strategy:** Has a plan in place for the mass extraction of all preserved objects

**Migration:** Converting the object from one format to another format considered to be of greater stability

**Monitoring:** Automatically checks for object corruption

**Auto Recovery:** Automatically overwrites the corrupted object with an uncorrupted redundant copy
Digital POWRR Workshop - Glossary of Digital Preservation Terms

<table>
<thead>
<tr>
<th>Term</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accession</td>
<td>The processes of receiving, preparing, cataloguing and storing digital resources in a form suitable for digital preservation. See also &quot;Deposit&quot; and &quot;Ingest.&quot;</td>
</tr>
<tr>
<td>AIP (Archival Information Package)</td>
<td>An Information Package, consisting of the Content Information and the associated Preservation Description Information, which is preserved within an OAIS.</td>
</tr>
<tr>
<td>Bag</td>
<td>A package of content that conforms to the BagIt Specification. Under the specification, a bag consists of a base directory containing a small amount of machine-readable text to help automate the content's receipt, storage and retrieval and a subdirectory that holds the content files. See also &quot;BagIt Specification&quot; and &quot;Bagger.&quot;</td>
</tr>
<tr>
<td>Bagger</td>
<td>A graphical software application tool to produce a package of data files that conforms to the BagIt Specification. See also &quot;Bag&quot; and &quot;BagIt Specification.&quot;</td>
</tr>
<tr>
<td>Checksum</td>
<td>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. See also &quot;Fixity Check.&quot;</td>
</tr>
<tr>
<td>Description</td>
<td>The process of capturing, analysing, organising and recording information that serves to identify, manage, locate and explain data resources and the contexts that produced them. See also &quot;Metadata.&quot;</td>
</tr>
<tr>
<td>Digital Preservation</td>
<td>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.</td>
</tr>
<tr>
<td>DIP (Dissemination Information Package)*</td>
<td>A version created from the original version of a data resource for dissemination to users. In OAIS terms, the Dissemination Information Package that the user receives, may be derived from one or more Archival Information Packages.</td>
</tr>
<tr>
<td>File Format</td>
<td>Packages of information that can be stored as data files consisting of a fixed byte-serialized encoding of a specified information model, and/or a fixed encoding of that encoding in a tangible form on a physical storage structure.</td>
</tr>
<tr>
<td>Fixity Check</td>
<td>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. See also &quot;Checksum&quot; and &quot;Digital Signature.&quot;</td>
</tr>
<tr>
<td>Format Migration</td>
<td>A means of overcoming technical obsolescence by preserving digital content in a succession of current formats or in the original format that is transformed into the current format for presentation. The purpose of format migration is to preserve the digital objects and to retain the ability for clients to retrieve, display, and otherwise use them in the face of constantly changing technology.</td>
</tr>
<tr>
<td>Ingest</td>
<td>The process through which digital objects are added into a managed environment.</td>
</tr>
<tr>
<td>Life Cycle</td>
<td>A set of iterative, modular processes that govern the creation, acquisition, selection, description, sustainability, access and preservation of digital content over time.</td>
</tr>
<tr>
<td>Metadata: Administrative</td>
<td>Administrative metadata comprises both technical and preservation metadata and is generally used for internal management of digital resources.</td>
</tr>
<tr>
<td>--------------------------</td>
<td>----------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Metadata: Descriptive</td>
<td>Metadata that identifies a resource and describes its intellectual content for purposes such as discovery and identification.</td>
</tr>
<tr>
<td>Metadata: Preservation</td>
<td>The contextual information necessary to carry out, document, and evaluate the processes that support the long-term retention and accessibility of digital content. Preservation metadata documents the technical processes associated with preservation, specifies rights management information, establishes the authenticity of digital content, and records the chain of custody and provenance for a digital object. See also &quot;PREMIS.&quot;</td>
</tr>
<tr>
<td>Metadata: Rights Management</td>
<td>Administrative metadata that indicates the copyrights, user restrictions, and license agreements that might constrain the end-use of digital content (including metadata files).</td>
</tr>
<tr>
<td>Metadata: Structural</td>
<td>Metadata used to describe the logical or physical types, versions, relationships or other characteristics of content files comprising a complex digital object.</td>
</tr>
<tr>
<td>Metadata: Technical</td>
<td>Metadata that describes the technical state of and process used to create a file. Often closely related either to its file format or the original software used to create the file, e.g. scanning equipment and settings used to create or modify a digital object.</td>
</tr>
<tr>
<td>MODS</td>
<td>Metadata Object Description Schema (MODS) is a schema for a bibliographic element set that may be used for a variety of purposes, and particularly for library applications. The standard is maintained by the Network Development and MARC Standards Office of the Library of Congress with input from users.</td>
</tr>
<tr>
<td>Open Archival Information System (OAIS) Reference Model</td>
<td>Developed by the Consultative Committee on Space Data, a conceptual framework and reference tool for defining a digital repository. It provides a model of the environment, functions, and data types for implementing a digital repository.</td>
</tr>
<tr>
<td>PREMIS</td>
<td>The international standard for metadata to support the preservation of digital objects and ensure their long-term usability. Developed by an international team of experts, PREMIS is implemented in digital preservation projects around the world, and support for PREMIS is incorporated into a number of commercial and open-source digital preservation tools and systems. The PREMIS Editorial Committee coordinates revisions and implementation of the standard, which consists of the Data Dictionary, an XML schema, and supporting documentation. See also &quot;Metadata: Preservation.&quot;</td>
</tr>
<tr>
<td>Preservation Copy</td>
<td>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.</td>
</tr>
<tr>
<td>SIP (Submission Information Package)</td>
<td>In the OAIS Reference Model, a SIP is an Information Package that is delivered by the producer to the repository for use in the construction of one or more ALPs. (See &quot;AIP&quot; and &quot;DIP&quot;)</td>
</tr>
</tbody>
</table>

Adapted from the NDAA Glossary ([http://www.digitaipreservation.gov/nnda/ndsaglossary.html](http://www.digitaipreservation.gov/nnda/ndsaglossary.html))
3-3-3 Action Plan

As a small group, identify existing institutional roles that you believe are important to involve in the process of creating a digital preservation program.

3 PEOPLE
Of all the roles identified by the group, list 3 individuals at your institution in these roles that you already have a working relationship with.

1. 
2. 
3. 

3 MONTHS
In the next 2 weeks, which of these individuals are you most comfortable reaching out to and discussing your collective next steps towards a digital preservation program?

After reaching out to this individual, who will be the next individual you both connect with in the next month?

Who will be the third individual that your burgeoning digital preservation advocacy team brings on board in the following month?

3 ACTIVITIES
List three activities you think your team can reasonably accomplish in your first three months together to move your program forward.

In 3 months, a member of our team will be emailing you a brief electronic survey to check the efficacy of the 3-3-3 Action Plan. We may also need to follow up with a phone call. Please provide us with the following information and accept our sincerest thanks for helping POWRR report its results to the Institute for Museum and Library Services.

Your Name: __________________________  Your Institution: __________________________
Your Email: __________________________  Your Phone #: __________________________
From Theory to Action

A pragmatic approach to digital preservation strategies and tools

A Digital POWRR Workshop

http://digitalpowrr.niu.edu/
Logistics/Housekeeping

- Basic Logistics

- Handouts/Flash Drives

- Assessment/Evaluation (today and in 3 months) YOUR FEEDBACK IS VITAL
  - Pre-Test
  - Post-Test
  - Standard Workshop Evaluation
  - 3 Month Follow-up

First Up... The Day’s Schedule!
## Morning

**Now – 9:45**

- Collect Pre-tests Expected Outcomes
- Who we are & How we got here
- Levels of Preservation (Activity)

**9:45 – 10:15**

- Solution in Theory vs. Solution in Practice
- Break

**10:15 – 11:00**

- Your Pre-Ingest Workflow
- Accessioning a Collection (Demo)

**11:00 – Noon**

- Tools that POWRR investigated

---

## Afternoon

**1:00 – 2:00**

- Solution in Action: Accessioning a Collection (Activity)

**2:00 – 2:30**

- Assembling Your Team Your 3-3-3 Action Plan (Activity)

**2:30 – Break**

**2:45 – 3:30**

- Advocacy, Policy, Potential Solutions Models

**3:30 – 3:50**

- Questions

**3:50 – 4:00**

- Post-Test
Expected Outcomes

- You will understand that different digital preservation tools/services can perform different functions within the digital curation lifecycle, and be able to explain how these tools/services can be used within your institution’s workflow.

- You will practice the initial pre-ingest steps necessary to accession a digital collection, as described in the OCLC report “Walk this Way,” and gain the skills necessary to repeat this process at your institution.

- You will gain hands on experience with a basic digital preservation tool and understand how it can be used within your institution’s workflow.

- You will take away resources that help align communication and advocacy, policymaking, and tool selection/implementation.

- You will create a 3-3-3 Action Plan to implement in the following 3 months that will move you closer to your digital preservation goals.
Who we are... and how we got here....

- Defining Moments → Found Some Friends
- Applied for an Implementation Grant → Received a “Figure It Out” Grant

We’ve learned a lot... and are a lot like you!

Proud to be works-in-progress:

[Logos of various universities]

Sponsored By: [Institute of Museum and Library Services]
Activity Time!
20 Minutes
NDSA Levels of Preservation

Where can my institution place its Bingo chips?

- We’ll go first
- Small Groups – Where do you think you fit in? (10 minutes)
- All Together – Poll of who is where!

NDSA: National Digital Stewardship Alliance
<table>
<thead>
<tr>
<th>Storage and Geographic Location</th>
<th>Level 1 (Protect your data)</th>
<th>Level 2 (Know your data)</th>
<th>Level 3 (Monitor your data)</th>
<th>Level 4 (Repair your data)</th>
</tr>
</thead>
<tbody>
<tr>
<td>- Two complete copies that are not collocated.</td>
<td>- At least three complete copies</td>
<td>- At least one copy in a geographic location with a different disaster threat.</td>
<td>- At least three copies in geographic locations with different disaster threats.</td>
<td></td>
</tr>
<tr>
<td>- For data on heterogeneous media (optical discs, hard drives, etc.) get the content off the medium and into your storage system.</td>
<td>- At least one copy in a different geographic location.</td>
<td>- Obsolescence monitoring process for your storage system(s) and media.</td>
<td>- Have a comprehensive plan in place that will keep files and metadata on currently accessible media or systems.</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>File Fidelity and Data Integrity</th>
<th>- Check fidelity on ingest if it has been provided with the content.</th>
<th>- Check fidelity on all ingests.</th>
<th>- Check fidelity of content at fixed intervals.</th>
<th>- Check fidelity of all content in response to specific events or activities.</th>
</tr>
</thead>
<tbody>
<tr>
<td>- Create fidelity metadata if it wasn't provided with the content.</td>
<td>- Use write-blockers when working with original media.</td>
<td>- Maintain logs of fidelity information; supply audit on demand.</td>
<td>- Ability to replace/repair corrupted data.</td>
<td></td>
</tr>
<tr>
<td>- Virus-check high risk content.</td>
<td>- Virus-check all content.</td>
<td>- Ability to detect corrupt data.</td>
<td>- Ensure no one person has write access to all copies.</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Information Security</th>
<th>- Identify who has read, write, move and delete authorization to individual files.</th>
<th>- Document access restrictions for content.</th>
<th>- Maintain logs of who performed what actions on files, including deletions and preservation actions.</th>
<th>- Perform audit of logs.</th>
</tr>
</thead>
<tbody>
<tr>
<td>- Restrict who has those authorizations to individual files.</td>
<td>- Store administrative metadata.</td>
<td>- Store standard technical and descriptive metadata.</td>
<td>- Store standard preservation metadata.</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Metadata</th>
<th>- Inventory of content and its storage location.</th>
<th>- Store transformative metadata and log events.</th>
<th>- Monitor file format obsolescence issues.</th>
<th>- Perform format migrations, emulation, and similar activities as needed.</th>
</tr>
</thead>
<tbody>
<tr>
<td>- Ensure backup and non-collocation of inventory.</td>
<td>- Monitor file formats in use.</td>
<td>- -</td>
<td>-</td>
<td></td>
</tr>
</tbody>
</table>

| File Formats | - When you can give input into the creation of digital files encourage use of a limited set of known open formats and codecs. | - Inventory of file formats in use. | - Monitor file format obsolescence issues. | - Perform format migrations, emulation, and similar activities as needed. |

http://www.digitalpreservation.gov/nds/activity/levels.html
So.....How do we get from here to there?

Solution in Theory

Solution in Practice

Scary OAIS Spaghetti Monster
Solution in Theory (riiiiiight)

- OAIS (Open Archival Information Systems) and other scary schematic models
- TRAC Certification (Trustworthy Repositories Audit & Certification)
- TDR ISO 16363 (Trustworthy Digital Repository ISO Standard)
- Complex Curation Lifecycles that don’t look a thing like your workflow
- ...and other things you don’t need to worry about when you
  just WANT TO GET STARTED and DO SOMETHING!

A Note: These are all valuable things that benefit the field of digital preservation greatly.... We just don’t want you to become overwhelmed by them and grind to a halt before you take your first steps...like we did!!
Solution in Practice
AKA Good Enough DP for real people!!

Our take on what you need to consider when thinking about your digital stuff.....
Our take on some things that need to happen or be considered along the way to this "Digital Preservation" thing....
Solution in Practice
AKA Good Enough DP for real people!!

Some things to keep in mind.....

• Not all tools and services are created equal.
  - Some tools/services do specific tasks (*microservices*).
  - Some tools/services combine multiple microservices
    *(you guessed it....*macroservices*!)*.

• Starting small is good enough!
  - Starting with a simple microservice tool will get you closer to your goals
    AND you can use them NOW!
  - Baby steps still move you forward..... See “Walk This Way”.

• Knowing what you have is crucial.
  - Write. It. Down. And maintain it.
  - We’ll show you an easy way to do this in a bit!
Let's Talk about Macroservices....

There are front-end/processing tools like.....
- Archivematica
- Curator's Workbench
- Data Accessioner

And there are back-end storage/preservation services like.....
- MetaArchive
- DuraCloud
- Amazon Glacier
- Internet Archive

There are even some services that will pretty much do it all like.....
- Preservica
- Dspace Direct (uses DuraCloud)
- Archivematica + DuraCloud

Note: Yes, there are also CMS's, IR software, and Forensics tools....ugh. However, these are outside the scope of this workshop!

AND you have to figure out what works best with what!!! But we have done some of that so you don't have to!
Clarification: Preservation vs. Access

Long term access (Preservation)

- **Purpose:** ensure long-term access
- **Focus:** current & future users
- Relies on *proven (reliable)* technologies to preserve digital objects across generations of technology
- *Accumulates* metadata over the life cycle to trace preserved content
- Preservation systems *create* new versions of digital objects for access as needs change over time

Short term access

- **Purpose:** provide content to users now
- **Focus:** current
- Relies on *cutting edge* technologies to provide best and fastest access at a point in time
- *Selects* metadata needed to use and understand content
- Access systems *deliver* objects with user-oriented services
BREAK TIME!
Back by 10:30, please

Next up: Your Pre-Ingest Workflow
Everyone’s favorite donor question:

Hey, do you want this jump drive for your collection/archive?

Demo: Accessioning a digital collection
Don't Panic - Your Pre-Ingest Workflow
aka Wrangling your digital stuff before you can get it into a shiny system

NOTE: This is only ONE way to do this... Everyone's workflow is a little different!

Starting from scratch:

- Begin an Inventory Spreadsheet
- Run accessioning tools (creates basic preservation metadata files in XML for you!)
  - Move everything to a stable carrier (like a network drive)
- Make an Access Copy from your Master Copy
- Continue populating Inventory Spreadsheet (if needed)
- OPTIONAL: Keep original media

✓ Most of these will cost you more time than money
✓ Document what you do pre-ingest. For future you.
✓ Remember: Good enough is just fine. For now.

Hang tight for a demo and some hands-on practice of this!
DIGITAL PRESERVATION DECISION FLOWCHART

Is this a unique collection of "enduring value" that was not commercially mass produced?

Yes

No

Do we have a contract to archive this collection?

Yes

Go to Digitization Contract Form

No

Check Do not archive content

Do you have permission from the copyright holder (if necessary) to archive this digital collection?

Yes

No

Check Do not archive content

Are any objects in this collection available in another Trusted Digital Repository?

Yes

Which Repository?

If collection is in another Trusted Digital Repository, check Do not archive content.

No

Are this collection available in a Trusted Print Repository or is there a hard copy available that will be kept long-term?

Yes

No

Check Full Preservation

Check Full Preservation

Is the hard copy deteriorating or in poor condition or on a near-obsolese format or media?

Yes

No

Check Full Preservation

Check Full Preservation

Would it be possible to recreate the digital collection from the hard copy?

Yes

No

Check Bit-Level Preservation

Check Do not archive content

DO NOT PRESERVE

FULL PRESERVATION

BIT-LEVEL PRESERVATION

Courtesy of:
Tawnya Keller, Digital Preservation Archivist
University of Utah
Pre-Ingest Inventory Spreadsheet Categories

These suggestions follow the recommended DPOE step “Identify” as locally defined by curator/archivist. Example at: http://www.carli.illinois.edu/sites/files/digital_collections/documentation/digipres_identify.pdf

☐ Category (digitization project; born digital; university archives)
☐ Title and Description
☐ Date(s) (date range of what’s IN there or date of creation if born digital)
☐ Location (CD, Jump drive, server location?)
☐ Extent (quantity: 48 journal issues; 106 images; 2 TB of video)
☐ Format (file formats: PDF, .Jpeg, Animated GIF, Wordstar2.0 file)

This is YOUR inventory... YOU get to decide if it needs additional fields, if some can be deleted, etc. You are the boss of this!

FILL OUT WHAT YOU CAN AS YOU WOULD WITH ANY NORMAL ACCESSION

DPOE is a Library of Congress Digital Preservation and Outreach Program
http://www.digitalpreservation.gov/education/
<table>
<thead>
<tr>
<th>Category</th>
<th>Title &amp; Description</th>
<th>Date</th>
<th>Location</th>
<th>Extent</th>
<th>Format</th>
</tr>
</thead>
<tbody>
<tr>
<td>(locally defined: project name? content creation method?)</td>
<td>(donor applied and/or yours...what's your practice?)</td>
<td>(YYYYMMDD or other locally defined format for accession date)</td>
<td>(storage place of choice--networked server recommended)</td>
<td>(quantity of folders, files by type or total size)</td>
<td>what extensions are involved: .jpg, .tif, .xls?</td>
</tr>
<tr>
<td>Special Collections, mixed: digitized and born digital</td>
<td>A Curator's Cat Collection. Donated by Jane (nee Pennypincher) and John Moneybags, Class of 2006. Feline Health research.</td>
<td>20140410</td>
<td>C:\Users\User\Desktop\NewAccessions\Masters</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Category</td>
<td>Feature</td>
<td>Cost</td>
<td>Other</td>
<td></td>
<td></td>
</tr>
<tr>
<td>--------------</td>
<td>------------------------------</td>
<td>--------</td>
<td>---------</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Clear Documentation</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Open Source</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Auto Recovery</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Monitoring</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Migration</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Exit Strategy</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Geographically Dispersed Data</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Redundancy</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Reliable, Long-Term Bit</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Storage</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Auto AIP Creation</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Auto DIP Creation</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Public Interface</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Auto SIP Creation</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Package Metadata</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Rights Management</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Manual Metadata</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Auto Metadata Harvest</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Auto Metadata Creation</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Auto Unique ID</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>File Dedupe</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Virus Scan</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Fidity Check</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Copy</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Digital DMRR</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Tool Evaluation Grid</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
1. Insert flash drive and open the explorer window

2. Navigate to DataAccessioner.jar and open it
Switch to live Data Accessioner demo
XML Output

Basic descriptive and Dublin Core metadata that you created

Extracted metadata: tree hierarchy of your accession (folder names, file names, last modified, size, and more!)

Fixity/Checksum: MD5

Identifying Information (Exiftool)

Identifying Information (file utility)
<table>
<thead>
<tr>
<th></th>
<th>Category</th>
<th>Title &amp; Description</th>
<th>Date</th>
<th>Location</th>
<th>Extent</th>
<th>Format</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>(locally defined: project name? content creation method?)</td>
<td>(donor applied and/or yours...what's your practice?)</td>
<td>YYYYMMDD or other locally defined format for accession date</td>
<td>(storage place of choice--networked server recommended)</td>
<td>(quantity of folders, files by type or total size)</td>
<td>what extensions are involved: .jpg, .tif, .xls?</td>
</tr>
<tr>
<td>2</td>
<td>Special Collections, mixed: digitized and born digital</td>
<td>A Curator's Cat Collection. Donated by Jane (nee Pennypincher) and John Moneybags, Class of 2006. Feline Health research. No restrictions on access; some material may have copyright restrictions by law.</td>
<td>20140410</td>
<td>C:\Users\User\Desktop\NewAccessions\Masters</td>
<td>42MB in four folders: Classic Kitties, Kitty Research, Kitty Videos, Stacey's Kitties</td>
<td>23 octet/stream, 10 pdf, 2 ppt, 24.jpg</td>
</tr>
</tbody>
</table>
Congratulations!

You just did the first few steps in the digital curation lifecycle.

<table>
<thead>
<tr>
<th>Digital POWRR Tool Evaluation Grid</th>
<th>Ingest</th>
<th>Processing</th>
<th>Access</th>
<th>Storage</th>
<th>Maintenance</th>
<th>Other</th>
</tr>
</thead>
<tbody>
<tr>
<td>Copy</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Fix/Check</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Virus Scan</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>File Dedupe</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Auto Unique ID</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Auto Metadata Creation</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Auto Metadata Harvest</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Manual Metadata</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rights Management</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Package Metadata</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Auto SIP Creation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Auto DIP Creation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Public Interface</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Auto DIP Creation</td>
<td></td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Auto AIP Creation</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>x</td>
<td></td>
</tr>
<tr>
<td>Migration</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Monitoring</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>Auto Recovery</td>
<td></td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Clear Documentation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Open Source</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cost</td>
<td>Free</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Well, we did it. *Your turn* comes after lunch!

But first..... Macroservices! WOO!
Macroservices: Doing it all! Sort of.

Using simple tools, like Data Accesioner, is what you can do while you are petitioning your institution for a more robust solution like...

- Archivematica
- Curator's Workbench
- DuraCloud
- MetaArchive
- Preservica
- Internet Archive

Please Keep In Mind...

This is NOT exhaustive

Software changes quickly!

Based on availability at time of testing and our perceived needs
Remember this?

Most tools and services only perform some of the functions in a digital curation lifecycle.

*Tools/Services in RED were tested in-depth by POWRR

Front-end * Processing

Archivematica
Curator's Workbench
Data Accessioner

Back-end * Storage * Preservation

MetaArchive
DuraCloud
Amazon Glacier
Internet Archive

Preservica
Dspace Direct (uses DuraCloud)
Archivematica + DuraCloud

But there are very few services that will pretty much do it all.
Front-end/Processing: Curator's Workbench

<table>
<thead>
<tr>
<th>Digital POWRR Tool Evaluation Grid</th>
<th>Ingest</th>
<th>Processing</th>
<th>Access</th>
<th>Storage</th>
<th>Maintenance</th>
<th>Other</th>
</tr>
</thead>
<tbody>
<tr>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Curator's Workbench: 
- **Copy**
- **Fixity Check**
- **Virus Scan**
- **File Dedupe**
- **Auto Unique ID**
- **Auto Metadata Creation**
- **Auto Metadata Harvest**
- **Manual Metadata**
- **Rights Management**
- **Package Metadata**
- **Auto SIP Creation**
- **Auto DIP Creation**
- **Auto AIP Creation**
- **Public Interface**
- **Exit Strategy**
- **Migration**
- **Monitoring**
- **Auto Recovery**
- **Clear Documentation**
- **Op...
Front-end/Processing: Curator’s Workbench

- Requires expertise in MODS & direct metadata entry
- Project partners couldn’t make it fit their workflows
- The update process is slow
- What we just demo’ed (DA) does a good enough version of this
# Front-end/Processing: Archivematica

<table>
<thead>
<tr>
<th>Digital POWRR</th>
<th>Ingest</th>
<th>Processing</th>
<th>Access</th>
<th>Storage</th>
<th>Maintenance</th>
<th>Other</th>
</tr>
</thead>
<tbody>
<tr>
<td>Archeumatica</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
</tbody>
</table>

---

**Archeumatica**
Front-end/Processing: Archivematica

- Open source/free software
- Requires IT support and administration (Virtual Machine, Ubuntu Server, etc.)
- Microservices run by themselves
- Shows all the steps for AIP, SIP, DIP
- Capability to upload own metadata
- Errors stop everything
- Great Google users group support
- Integrates with Content DM & DSpace
- Bundled with ICA-AToM (archival content management system like ARCHON)
- Hosted version now available
- File transfers not intuitive
- Slower processing, but that could be due to the fact that we are used to desktop-based applications
Archivematica: Transfer Collection

![Archivematica Transfer Interface](image)

<table>
<thead>
<tr>
<th>Transfer</th>
<th>UUID</th>
<th>Transfer start time</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample series</td>
<td>89a46845-0bed-4917-a482-ea004a798b9d</td>
<td>2013-10-10 13:06</td>
</tr>
</tbody>
</table>

- **Job:** Create SIP(s) [?]  
- **Job:** Load options to create SIPs  
- **Job:** Check transfer directory for objects  
- **Micro-service:** Complete transfer  
- **Micro-service:** Characterize and extract metadata  
- **Job:** Load labels from metadata/file_labels.csv  
- **Job:** Characterize and extract metadata  
- **Job:** Identify file format  
- **Job:** Determine which files to identify  
- **Job:** Select file format identification command  
- **Job:** Move to select file ID tool  
- **Micro-service:** Clean up names  
- **Micro-service:** Scan for viruses

Actions:
- **Awaiting decision**
- **Actions**
  - Create SIP(s) manually
  - Send to backlog
  - Reject transfer
  - Create single SIP and continue processing
Archivematica: Normalization On Ingest

<table>
<thead>
<tr>
<th>Submission Information Package</th>
<th>UUID</th>
<th>Ingest start time</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample_series</td>
<td>2c5fe6dbf-202-4932-8f8c-1063e5579dd</td>
<td>2013-10-10 13:13</td>
</tr>
</tbody>
</table>

**Micro-service: Normalize**

- Job: Normalize
- Job: Resume after normalization file identification tool selected.
- Job: Identify file format
- Job: Select pre-normalize file format identification command
- Job: Move to select file ID tool
- Job: Set resume link after tool selected.
- Job: Find options to normalize as
- Job: Move to workflowDecisions-createDIP directory
- Job: Grant normalization options for no pre-existing DIP
- Job: Set remove preservation and access normalized files to renormalize link.
- Job: Check for Access directory
- Job: Check for Service directory
- Job: Identify manually normalized files

**Actions**

- Normalization for preservation and access
- Normalization for preservation
- Reject SIP
- Normalize service files for access
- Do not normalize
- Normalize manually
- Normalize for access

**Status**

- Completed successfully
- Completed successfully
- Completed successfully
- Completed successfully
- Completed successfully
- Completed successfully
- Completed successfully
- Completed successfully
Archivematica: Add AIP to Storage
Back-end/Preservation: DuraCloud

<table>
<thead>
<tr>
<th>Digital POWRR Tool Evaluation Grid</th>
<th>Ingest</th>
<th>Processing</th>
<th>Access</th>
<th>Storage</th>
<th>Maintenance</th>
<th>Other</th>
</tr>
</thead>
<tbody>
<tr>
<td>DuraCloud</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
</tbody>
</table>
Back-end/Preservation: DuraCloud

- Nonprofit; Open Pricing; Community buy-in
- Cloud storage/preservation solution
- Different storage provider options
- Hosted service (requires little to no IT support on your end!)
- Some microservices available (like health checks that verify checksums)
- Different options/methods for uploading content (bulk, single item, etc.)
- Intuitive uploads and file management
- Easy exit strategy
- Easy integration with DSpace
- New: Integrated with hosted version of DSpace
- Media streaming based on Amazon’s Cloud service
- Responsive customer service with very good documentation
- Affordable; Scalable; Easy to get started
DuraCloud.org
Head to the website for...

- Open Pricing
- Free Trial
- Lots of webinars and tutorials
- Learn more about the new DSpace Direct... a hosted version of the DSpace Institutional Repository software that is integrated with DuraCloud for preservation

<table>
<thead>
<tr>
<th>DuraCloud Preservation</th>
<th>DuraCloud Preservation Plus</th>
<th>DuraCloud Enterprise</th>
<th>DuraCloud Enterprise Plus</th>
</tr>
</thead>
<tbody>
<tr>
<td>The DuraCloud Preservation plan is ideal for institutions that wish to store one copy of their content in the cloud. Subscription plan is available with storage between 1-5TB of content. Example use case: Back-up preservation storage for a small amount of content.</td>
<td>The DuraCloud Preservation Plus plan is best suited for organizations that wish to store two copies of their content in the cloud. Subscription plan is available with storage between 1-5TB of content. Example use case: Archival storage for a moderate amount of content.</td>
<td>The DuraCloud Enterprise plan is designed to meet the needs of institutions that wish to store one copy of their content in the cloud and need to provide a variety of individuals, departments, research groups, etc. access to a single DuraCloud account. Subscription plan is available with unlimited storage. Contact us for custom quote for storage beyond 10TB. Example use case: Long-term access storage for a variety of institutional content.</td>
<td>The DuraCloud Enterprise Plus plan is intended to assist organizations that wish to store two copies of their content in the cloud and need to provide a variety of individuals.</td>
</tr>
<tr>
<td>Standard features</td>
<td>Standard features</td>
<td>Standard features</td>
<td>Standard features</td>
</tr>
<tr>
<td>- Automatic synchronization of content between primary and secondary storage providers</td>
<td>- Choice of secondary cloud storage providers</td>
<td>- Media serving</td>
<td>- Automatic synchronization of content between primary and secondary storage providers</td>
</tr>
<tr>
<td>- Choice of secondary cloud storage providers</td>
<td>- Automatic file recovery between copies</td>
<td>- Account management</td>
<td>- Choice of secondary cloud storage providers</td>
</tr>
<tr>
<td>- Automatic file recovery between copies</td>
<td>- Storage in Amazon S3:</td>
<td>- Storage in Amazon S3 + Amazon Glacier:</td>
<td>- Storage in Amazon S3 + Amazon Glacier:</td>
</tr>
<tr>
<td></td>
<td>$1,806 (subscription which includes 1TB storage)</td>
<td>$1,925 (subscription which includes 1TB storage)</td>
<td>$5,500 (subscription which includes 1TB storage)</td>
</tr>
<tr>
<td></td>
<td>$700 for additional TBs</td>
<td>$825 for additional TBs</td>
<td>$500 for additional TBs</td>
</tr>
<tr>
<td></td>
<td>$2,800 (subscription which includes 1TB storage)</td>
<td>$1,400 for additional TBs</td>
<td>$5,625 (subscription which includes 1TB storage)</td>
</tr>
<tr>
<td></td>
<td>$625 for additional TBs</td>
<td></td>
<td>$625 for additional TBs</td>
</tr>
</tbody>
</table>
## Back-end/Preservation: MetaArchive

<table>
<thead>
<tr>
<th>Digital POWRR Tool Evaluation Grid</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>MetaArchive (A private LOCKSS Network)</strong></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Ingest</th>
<th>Processing</th>
<th>Access</th>
<th>Storage</th>
<th>Maintenance</th>
<th>Other</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
</tbody>
</table>
Back-end/Preservation: MetaArchive

- Nonprofit; Open Pricing
- Instant community in the Cooperative! - All the cool kids are doing it!
- Helpful and responsive customer service
- Private LOCKSS network
- Dark Archive
- Requires dedicated IT administration
- Most memberships require attending meetings

- Assumes pre-processing work is done
- Rules for minimum processing requirements (ie file naming conventions)

Sample costs [http://www.metaarchive.org/costs](http://www.metaarchive.org/costs)
If an institution wanted to preserve 2 TB of content with us, they would pay:

Sustaining Member: [$5,500 (membership) + $2,000 (space) x 3 years] + $4,600 (server) = $27,100/3 years, or $9,033/year

Preservation Member: [$3,000 (membership) + $2,000 (space) x 3 years] + $4,600 (server) = $19,600/3 years, or $6,533/year

Collaborative Member (mock example comprised of 20 institutions): [$4,000 (membership) + $2,000 (space) x 3 years] + $4,600 (server) = $22,600/3 years or $7,533/year total or $377.00/year per member
Back-end/Preservation: MetaArchive

Collaborative Membership Model

MetaArchive

Lead Institution

Staging Server

LOCKSS Server

Partner 1

Partner n

Partner 2

AIPs

Tech Help

AIPs via FTP

Very simplified version of how it works:

1) Partners prepare their content for preservation and package it.
   → We used the BagIt specification, and Bagger helped us with this

2) Lead Institution prepares a staging server, sets appropriate access protocols and assists Partners with technical help.

3) Partners FTP their AIP’s (Bags) to the staging server at the Lead Institution.
   → We used Filezilla

4) MetaArchive harvests the AIP’s from the Lead Institution’s staging server and pushes it into their LOCKSS network.

One other thing: The Lead Institution also has a dedicated server that runs the LOCKKS software, is hooked into the MetaArchive network of servers across the globe, and is actively preserving the content of other Members.
Front-end & Back-end: Preservica

<table>
<thead>
<tr>
<th>Digital POWRR Tool Evaluation Grid</th>
<th>Ingest</th>
<th>Processing</th>
<th>Access</th>
<th>Storage</th>
<th>Maintenance</th>
<th>Other</th>
<th>Cost</th>
</tr>
</thead>
<tbody>
<tr>
<td>Preservica(Tessella)</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
</tbody>
</table>

- **Copy**
- **Fixity Check**
- **Virus Scan**
- **File Dedupe**
- **Auto Unique ID**
- **Auto Metadata Creation**
- **Auto Metadata Harvest**
- **Manual Metadata**
- **Rights Management**
- **Package Metadata**
- **Auto SIF Creation**
- **Auto AIP Creation**
- **Auto DIP Creation**
- **Public Interface**
- **Auto Exit Creation**
- **Redundancy**
- **Redundancy Exit Protection**
- **Geography**
- **Strategy**
- **Migration**
- **Monitoring**
- **Auto Recovery**
- **Open Source**
- **Clear Documentation**

- **Cost**: Varies
Front-end & Back-end: Preservica

- All encompassing:
  - Ingest
  - Processing
  - End-User Access
  - Preservation
  - Migration

- Aligned with OAIS reference model

- Hosted Service (Requires little IT support on your end)

- Very user friendly

- Intuitive workflows

- Exit strategy available (batch export)

- Ability to harvest via web crawls

- Solid customer support

- Different training options available for institutions with smaller budgets

- Currently uses only Amazon cloud storage
  - new options forthcoming

- Proprietary, vendor-based
<table>
<thead>
<tr>
<th>Plan Description</th>
<th>Storage Capacity</th>
<th>Price per Year</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ideal for smaller archives and collections that anticipate slow growth in content.</td>
<td>up to 100 GB in Amazon S3</td>
<td>$3,950 per year</td>
</tr>
<tr>
<td>Ideal for smaller archives and collections that anticipate medium growth in content.</td>
<td>up to 250 GB in Amazon S3</td>
<td>$6,950 per year</td>
</tr>
<tr>
<td>Ideal for small to mid-size archives and collections that anticipate medium growth in content.</td>
<td>up to 500 GB in Amazon S3</td>
<td>$8,950 per year</td>
</tr>
<tr>
<td>Ideal for larger archives that anticipate medium growth and need fast access to content in S3 and/or lower cost Glacier storage for infrequently accessed content.</td>
<td>1 to 10 TB in Amazon S3 and/or Amazon Glacier</td>
<td>$11,950 per year for first TB in S3, then $1,950 per TB S3 per year and/or $750 per TB Glacier per year</td>
</tr>
</tbody>
</table>

**Standard features for all plans**

- Full suite of OAIS compliant workflows
- Public access/discovery module
- Active Preservation technology
- Automated DSpace ingest
- CALM catalogue synchronization
- Linked Data Registries
- All hosting, backup and operations
- Professional software maintenance and upgrades
- Unlimited support by email
- One day introductory training on our regular webinars
- Metadata held in fast Amazon RDS storage
- Durable Cloud storage - multiple copies, multiple locations, integrity checking
- Amazon S3 and/or Amazon Glacier (above 1 TB)
- Copy back to local server option
- No penalty upgrades to higher plans
- No cost to retrieve content if you decide to leave the service
<table>
<thead>
<tr>
<th>State</th>
<th>Name</th>
<th>Progress</th>
<th>Started</th>
<th>Finished</th>
<th>Messages</th>
</tr>
</thead>
<tbody>
<tr>
<td>✓</td>
<td>Import from Transfer Area</td>
<td></td>
<td>07.03.14 14:33:09</td>
<td>07.03.14 14:33:14</td>
<td></td>
</tr>
<tr>
<td>✓</td>
<td>Virus Check</td>
<td></td>
<td>07.03.14 14:33:14</td>
<td>07.03.14 14:33:47</td>
<td></td>
</tr>
<tr>
<td>✓</td>
<td>Fixity Check</td>
<td></td>
<td>07.03.14 14:33:47</td>
<td>07.03.14 14:33:50</td>
<td></td>
</tr>
<tr>
<td>✓</td>
<td>Metadata Integrity</td>
<td></td>
<td>07.03.14 14:33:50</td>
<td>07.03.14 14:33:53</td>
<td></td>
</tr>
<tr>
<td>✓</td>
<td>Content Integrity</td>
<td></td>
<td>07.03.14 14:33:53</td>
<td>07.03.14 14:33:56</td>
<td></td>
</tr>
<tr>
<td>✓</td>
<td>SIP Validation</td>
<td></td>
<td>07.03.14 14:33:56</td>
<td>07.03.14 14:33:59</td>
<td></td>
</tr>
<tr>
<td>✓</td>
<td>SIP Validation with Database Crosscheck</td>
<td></td>
<td>07.03.14 14:33:59</td>
<td>07.03.14 14:34:02</td>
<td></td>
</tr>
<tr>
<td>✓</td>
<td>Characterise</td>
<td></td>
<td>07.03.14 14:34:02</td>
<td>07.03.14 14:34:17</td>
<td>View</td>
</tr>
<tr>
<td>⚪</td>
<td>Store Files</td>
<td></td>
<td>07.03.14 14:34:17</td>
<td>08.03.14 05:09:08</td>
<td>View</td>
</tr>
</tbody>
</table>
Preservation Plan Details

<table>
<thead>
<tr>
<th>Transformation Name</th>
<th>9605e41e-2959-4c0f-b6a1-73fd0b6aad97</th>
</tr>
</thead>
<tbody>
<tr>
<td>Description</td>
<td>Test</td>
</tr>
<tr>
<td>Date Created</td>
<td>1</td>
</tr>
<tr>
<td>Number Of File Sets</td>
<td>1</td>
</tr>
<tr>
<td>Migration Pathway Role</td>
<td>Preservation</td>
</tr>
<tr>
<td>Usage</td>
<td>Test</td>
</tr>
<tr>
<td>Top-Level Collection</td>
<td>Danielle</td>
</tr>
</tbody>
</table>

Exit | Terminate

This transformation is ready for execution.

Selected File Sets

<table>
<thead>
<tr>
<th>Collection Title</th>
<th>Catalogue Ref</th>
<th>Description</th>
<th>Manifestation</th>
<th>Number of Files</th>
<th>Files at Risk</th>
<th>Size</th>
<th>Size of Files at Risk</th>
<th>Details</th>
</tr>
</thead>
<tbody>
<tr>
<td>Danielle</td>
<td>RC 278 Kishwaukee Kiwanis</td>
<td>RC 278 Kishwaukee Kiwanis</td>
<td>1</td>
<td>31</td>
<td>2</td>
<td>15.41 MB</td>
<td>592.03 kB</td>
<td>Details</td>
</tr>
<tr>
<td>Totals</td>
<td>1</td>
<td></td>
<td></td>
<td>31</td>
<td>2</td>
<td>15.41 MB</td>
<td>592.03 kB</td>
<td></td>
</tr>
</tbody>
</table>

Selected Migration Pathways

<table>
<thead>
<tr>
<th>Source Format</th>
<th>Version</th>
<th>PUID</th>
<th>Target Format</th>
<th>Version</th>
<th>PUID</th>
<th>Tool</th>
</tr>
</thead>
<tbody>
<tr>
<td>Microsoft Word for Windows</td>
<td>2007 onwards</td>
<td>fmt/412</td>
<td>Acrobat PDF 1.4 - Portable Document Format</td>
<td>1.4</td>
<td>fmt/18</td>
<td>Open Office PDF</td>
</tr>
</tbody>
</table>

Back | Confirm
### Workflow Details

<table>
<thead>
<tr>
<th>Workflow Context</th>
<th>Migration (Filtered)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Workflow Definition</td>
<td>Preservation Workflow (Filter by Top-Level Collection)</td>
</tr>
<tr>
<td>Workflow ID</td>
<td>879</td>
</tr>
<tr>
<td>Workflow State</td>
<td>Active</td>
</tr>
<tr>
<td>Date Started</td>
<td>13.04.14 15:47:58</td>
</tr>
<tr>
<td>Date Finished</td>
<td></td>
</tr>
<tr>
<td>Number Of File Sets</td>
<td>1</td>
</tr>
<tr>
<td>Migration Pathway Role</td>
<td>Preservation</td>
</tr>
</tbody>
</table>

#### Step Progress

<table>
<thead>
<tr>
<th>State</th>
<th>Name</th>
<th>Progress</th>
<th>Started</th>
<th>Finished</th>
<th>Messages</th>
</tr>
</thead>
<tbody>
<tr>
<td>🔄</td>
<td>Migrate AIPs</td>
<td></td>
<td>13.04.14 15:59:59</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Step Progress

<table>
<thead>
<tr>
<th>State</th>
<th>Progress</th>
<th>Started</th>
<th>Finished</th>
<th>Messages</th>
</tr>
</thead>
</table>

Preservation Plan Details

<table>
<thead>
<tr>
<th>Transformation Name</th>
<th>9605ea1e-2959-4c0f-b6a1-73fd0b6aad97</th>
</tr>
</thead>
<tbody>
<tr>
<td>Description</td>
<td>Test1</td>
</tr>
<tr>
<td>Number Of File Sets</td>
<td>1</td>
</tr>
<tr>
<td>Migration Pathway Role</td>
<td>Preservation</td>
</tr>
<tr>
<td>Usage</td>
<td>Test</td>
</tr>
</tbody>
</table>

The transformation has completed.

<table>
<thead>
<tr>
<th>Top Level Record</th>
<th>Collection Title</th>
<th>Outcome</th>
</tr>
</thead>
<tbody>
<tr>
<td>RC 278 Kishwaukee Kiwanis</td>
<td>Danielle</td>
<td>Transformation Complete</td>
</tr>
</tbody>
</table>
Front-end & Back-end: Internet Archive

<table>
<thead>
<tr>
<th>Digital POWRR</th>
<th>Ingest</th>
<th>Processing</th>
<th>Access</th>
<th>Storage</th>
<th>Maintenance</th>
<th>Other</th>
</tr>
</thead>
</table>

Internet Archive   | x  | x  | x  | x  | x  | x  | x  | x  | x  | x  | x  | x  | x  | x  | x  | x  | x  | x  | x  | x  | x  | x  | x  | x  | x  | x  | x  | x  | x  | x  | Free |
Internet Archive

- Only intended for materials in the public domain (available to everyone).
- Geographically distributed copies.
- No frills (and no charge!) service.

Share your Files

Please contribute books, audio, and video files that you have the right to share. The Internet Archive, a non-profit library, will provide free storage and access to them. (We reserve the right to remove any submitted material.)

Upload Files

Or choose an alternate upload method:

- Classic Uploader
- Live Music Archive
- FTP

Handle books best, but can accommodate manuscripts, audio, video, and images.

Is especially suited for small (VERY small institutions with limited (or no) other alternatives.

Does offer a more traditional preservation product through its Archive-It service.
Activity: Accessioning a Digital Collection 1 - 2pm

1. Insert flash drive and open the explorer window
   Data Accessioner
   Donated Collection Folder
   Digital Collections Inventory file
   Other stuff.....

2. Navigate to DataAccessioner.jar and open it
Create your accession directory:
Where you want the collection to go live
Preferably a stable media like your network drive

In your POWRR drive, open the **NewAccessions** folder and select the **Master Copies** folder.
Select the collection you are accessioning
Populate descriptive metadata and migrate your collection

<table>
<thead>
<tr>
<th>Date</th>
<th>City</th>
<th>Event Name</th>
<th>Contact</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Apr 9, 2014</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Apr 9, 2014</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Apr 9, 2014</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Apr 9, 2014</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

You will be able to see the progress bar move at the bottom.

Select which element you want to add metadata to.

Add the Dublin Core Metadata goes here.

Hit the "Migrate" button to begin the migration process.
What did you create?

New copy of your migrated collection.

Located in the Directory that you specified.

XML Metadata file

You. Are. AWESOME.
Not quite finished...

Make a copy of the Master, place in the Access Copies folder, and don’t touch the Master Copy again unless a new derivative is needed or until you move it into a preservation system!!
Not quite finished...

And finally...update your Inventory to reflect the location of the Access Copy. Note addition of XML file after processing.

<table>
<thead>
<tr>
<th>Category</th>
<th>Title &amp; Description</th>
<th>Date</th>
<th>Location</th>
<th>Extent</th>
<th>Format</th>
</tr>
</thead>
<tbody>
<tr>
<td>Special Collections, mixed:</td>
<td>A Curator's Cat Collection. Donated by Jane (nee Pennypinchener) and John Moneybags, Class of 2006. Feline Health research.</td>
<td>20140410</td>
<td>C:\Users\User\Desktop\NewAccessions\Masters</td>
<td>42MB in four folders: Classic Kittles, Kitty Research, Kitty Videos, Stacey's Kittles</td>
<td>23 octet/stream, 10 pdf, 2 ppt, 24 .jpg, 1 XML file from DDA</td>
</tr>
<tr>
<td>digitized and born digital</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Special Collections, mixed:</td>
<td>A Curator's Cat Collection. Donated by Jane (nee Pennypinchener) and John Moneybags, Class of 2006. Feline Health research.</td>
<td>20140410</td>
<td>C:\Users\User\Desktop\NewAccessions\AccessCopies</td>
<td>42MB in four folders: Classic Kittles, Kitty Research, Kitty Videos, Stacey's Kittles</td>
<td>23 octet/stream, 10 pdf, 2 ppt, 24 .jpg, 1 XML file from DDA</td>
</tr>
<tr>
<td>digitized and born digital</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
CONGRATULATIONS!

We call this "Digital Preservation in Your Office"

There are things that need to happen outside of your office as well....
Outside Your Office

Digital Preservation is not sustainable by just using a tool or selecting a service. Sustainability takes funding and people.

You cannot do this alone. You will need to talk to other people... because you are not the only boss of this.

Successful Digital Preservation programs take a team of people at multiple administrative levels.

Three-Legged Stool of Digital Preservation

Anne R. Kenney
Nancy McGovern
Digital Preservation Management Workshop
http://www.dpworkshop.org/
Assemble Your Team!

Image: Flickr Commons
Outside Your Office

Group Activity: 3-3-3 Action Plan

Create a list of all roles in an organization that should play a part in some aspect of digital preservation
3-3-3 Action Plan: Build Your Team

Now let’s move from roles to people....

- On your 3-3-3 Action Plan handout, list 3 individuals at your institution in these roles that you already have a working relationship with.

- Which of these folks are you willing to contact in the next 2 weeks?
  ...in the following month?
  ...in the following 3 months?

- After bringing these colleagues on board, what are 3 concrete, small steps that you can take together to move your burgeoning DP program forward?

~Conversations/Meetings ~Inventory what you already have
~Enhance the metadata of the records you already have
~Look at how current policies address digital materials (ex. collection development)
~Tool investigation: Dig a little deeper on tools that piqued your interest today
  ~Look at other institutions’ DP policies with an eye to crafting your own
  ~Engage in some outreach/education activities...host a Brown Bag!
~Download DDA and play with it some more! ~Read the POWRR white paper
Now Let’s Assess...

How will you know if your 3 activities succeeded?

- Added people to team?
- Number of people newly educated?
- Number of items added to inventory?
- Number of tools investigated?
- Number of DP policies looked at?
- Revised standing policies?
BREAK TIME!
Back by 2:45, please

Next Steps:
Advocacy, Policy, and Potential Solution Models
Next Steps: Advocacy

- Advocacy is valuable because you’re educating people about why digital preservation is also THEIR problem.
  - Our one-pagers may help you frame why digital preservation is important to different jobs/function.
  - The risks of doing nothing are a lot greater than they may think.

- Good policies incorporate multiple viewpoints.

- Other people at your institutions will bring up issues – and possible solutions – you may have missed.

- You will discover many things that you don’t directly control that still directly affect your work. This will lead you to more people to add to your team.
Next Steps: Towards a Policy

You have started assembling your team....now what?

We found a gap analysis *really* helpful:

- Where are you now?
- Where would you ideally like to be?
- What is keeping your institution from moving in that direction?
- What are some interim steps you can take to move in the right direction?
Next Steps: Towards a Policy

We also found that Gap Analyses can be challenging...

- Be brutally honest. It’s the only way to move forward.
- Look closely at risk: What is the cost of doing nothing?
- Documenting what you know will tell you what you don’t know.
- Feel free to look at our case studies and see how it worked. Our wiki has the case studies of all 5 of the POWRR partner institutions.

Next Steps: Towards a Policy

- Policy is where the ideal leads to actual workflows within your organization.

- We’ve collected links to resources on our website for getting started.  
  http://digitalpowrr.niu.edu/digital-preservation-101/

- Iterative is TOTALLY OKAY.

- Begin with how you would LIKE your workflow to run: tool selection may come out of that.

- Does it work with your already existing policies?

- It’s better to have a draft policy ready if resources show up than random resources with no policy.

- DOCUMENT what you do….future you will be deeply grateful.
Next Steps: Potential Solution Models
How to Decide? Results May Vary...

Things to consider:

- How many staff members will be actively engaged in the digital curation lifecycle? Are they tech-savvy?

- How robust and supportive is your technical/systems group? Do you even have one? How about some developers/programmers...have any of those on staff?

- Is your institution already using archival management software or an Institutional Repository (like ARCHON/ArchivesSpace, BePress, Fedora etc.)? You'll want to select tools/services that work well with what you have.

- Do you have digital collections unique to your institution that are irreplaceable? Consider organizing collections along the lines of those that warrant more robust preservation services than others. For example:

  | 1 TB (High Value) | MetaArchive (gold standard) |
  | 3 TB (Medium Value) | Amazon Glacier (cheapest storage with fixity checking) |
  | Rest (Replaceable) | Tape Drive Backups |

In other words: One tool/service will not be your only solution.
How to Decide? Results May Vary...
Remember: Smaller institutions with less resources may also have unique advantages like….

- Less red tape for getting things done
- Fewer levels to push requests for additional resources through
- Self-administered workstations (aka no IT administrative lock downs)
- Personnel-heavy operating model (usually has smaller cash flow)
- Higher cash flows and less data (like small, private institution)

It doesn't take years to set up an account with something like DuraCloud.
You only need to convince the person one level above you to get what you need.
Want to install a simple open source tool? Go for it!

This is ideal for running a *free* robust tool that requires a developer and server administrator like Archivematica.
You can purchase a reasonably-priced, hosted soup-to-nuts solution.
Next Steps: Potential Solution Models

POWRR White Paper available at:
http://commons.lib.niu.edu/handle/10843/13610
Wrapping Up

Our Final Thoughts & Your Questions

I survived the POWRR workshop! Now what?

https://digitalPOWRR.niu.edu/survived-powrr-wkshp/

We’re here to help. Seriously.

YOU CAN DO THIS. Really. But not alone. So bring some friends.

“If you want to go fast...go alone. If you want to go far...go together.” — African Proverb

Remember: Baby steps still move you forward!
Evaluation Time! (10 minutes)

- Post-Test
- Workshop evaluation...tells us about pace, style of presenting, etc.

In 3 Months...

- Emailing you a brief survey around your 3-3-3 Action Plan
- There may be a phone call too (if we don’t hear back from you or if we need further details)

Please note: The IMLS requires us to do these things...and it helps to make sure these workshops are delivering outcomes that bring tangible results to our peers!
Thank You for Coming!

Please Return:
- Jump Drives
- Contact Info Section of 3-3-3 Action Plan (bottom)
- Pre & Post Tests
- Workshop Evaluation
POWRR Project Team Members
Contact us...we are here to help!

Northern Illinois University
*Lynne M. Thomas
Imlthomas@niu.edu
Curator, RBSC
815.753.0255

Drew VandeCreek
drew@niu.edu
Director Digital Scholarship
815.753.7179

*Jaime Schumacher
jschumacher@niu.edu
Digital POWRR Director
815.753.0576

Stacey Erdman
serdman@niu.edu
Digital Collections Curator
815.753.1004

*Danielle Spalenka
dspalenka@niu.edu
Regional Hist Cntr Curator
815.753.9394

Matthew Short
mshort@niu.edu
Metadata Librarian
815.753.9868

Nathan Books
nbooks@niu.edu
Technical Associate
815.753.9653

Chicago State University

Aaisha Haykal
ahaykal@csu.edu
University Archivist
773.995.3843

Martin Kong
mkong@csu.edu
Systems Librarian
773.995.3908

Illinois State University

Patrice-Andre Prud’homme
ppprudh@ilstu.edu
Digital Collections Head
309.438.5385

Illinois Wesleyan University

Meg Miner
mminer@iwu.edu
University Archivist
309.556.1538

Western Illinois University

*Jeff Hancks
jl-hancks@wiu.edu
Director, Archives and Special Collections
309.298.2717

* Presented Today
Using Digital POWRR DP One Pagers

It can be difficult to explain the importance of digital preservation to various groups of people because they may not share the same vocabulary and viewpoint on the topic. These one-pagers are designed to target various audiences and educate them on the topic. They are meant to show the specific audience why digital preservation should be important to them and how not doing something about the problem may affect them.

These one-page handouts are also just a starting point. If you would like to use them “out of the box” you are welcome to do so. The only edits you will need to make are adding your contact information in the right column and deleting the header that designates the intended audience. However, you are free to edit the points to more specifically target an audience and more accurately reflect the current digital preservation situation at your institution.

If you have any further questions or would like more information about the Digital POWRR project please visit digitalpowrr.niu.edu.

This project is made possible by a grant funded by the Institute of Museum and Library Services.

The Institute of Museum and Library Services is the primary source of federal support for the nation's 123,000 libraries and 17,500 museums. Through grant making, policy development, and research, IMLS helps communities and individuals thrive through broad public access to knowledge, cultural heritage, and lifelong learning.
Why Should Digital Preservation Matter to You?

Everyone has a role to play in safeguarding our university’s digital assets.

The intellectual property of our institution is at risk. We must have an active digital preservation program in place to be a leading institution. Without such a program all the content institution members create is at risk of being lost.

The landscape for how research is performed has changed dramatically. Current estimates are that without digital preservation electronic records could become inaccessible within a few years, a decade at most. To ensure long-term viability and relevance of our work we need to begin a digital preservation program now as a core component of campus activities like teaching or research.

Facts about Digital Preservation

- Faculty, students and staff are creating valuable content daily (e.g., research data, photo archives, assessment data, and teaching materials/syllabi). Without a program in place to ensure the viability of that content over the long-term, those files will become unreadable.

- An Institutional Repository makes digital content available and accessible, but a digital preservation program needs to be in place to maintain those files.

- A robust Institutional Repository program with long-term digital preservation attracts faculty and students by showcasing and preserving faculty work.

- Appropriately managing research data to prevent catastrophic data loss is a critical issue for faculty. A catastrophic data loss could prevent faculty from completing their work to get tenure. Failure to achieve tenure causes faculty turnover, which is costly to the institution.

- Digital preservation has varying levels of affordability but some form of it is necessary for every-size institution. It can be implemented incrementally, with greater benefits achieved through higher levels of implementation.

What is Digital Preservation?

Digital preservation combines policies, strategies and actions to ensure long-term access to content that is born digital or converted to digital form regardless of the challenges of file corruption, media failure and technological change.

Digital preservation is an ongoing process, not a one-time activity, and needs to be addressed programmatically. Digital Preservation is more than a back-up.

Why do we need it?

Digital files become inaccessible over time unless they are managed and migrated to new technology as needed. Even content that resides on a web site or social media site is at risk.
Why Should Digital Preservation Matter to You?

Everyone has a role to play in safeguarding our university’s digital assets.

The intellectual property of our institution is at risk. We must have an active digital preservation program in place to be a leading institution. Without such a program all the content institution members create is at risk of being lost.

The landscape for how institutional memory and culture are created has changed dramatically. Current estimates are that without digital preservation electronic records could become inaccessible within a few years, a decade at most. To ensure long-term viability and relevance of our work we need to begin a digital preservation program now as a core component of campus activities like teaching or research.

---

**Facts about Digital Preservation**

➢ Faculty, students and staff are creating valuable content daily (e.g., photo and other media archives, works of artistic and scholarly expression). Without a program in place to ensure the viability of that content over the long-term, those files will become unreadable.

➢ An Institutional Repository makes digital content available and accessible, but a digital preservation program needs to be in place to maintain those files as technology changes.

➢ We compete with institutions where digital preservation is ongoing, potentially putting us at a disadvantage when recruiting. Implementing a digital preservation program will give us a competitive advantage over institutions where projects and programs are not yet in place.

➢ Digital preservation isn’t just about judicious backing up of files. We need to preserve these works in order to attract students by showcasing and preserving our community’s work.

---

**What is Digital Preservation?**

Digital preservation combines policies, strategies and actions to ensure long-term access to content that is born digital or converted to digital form regardless of the challenges of file corruption, media failure and technological change.

Digital preservation is an ongoing process, not a one-time activity, and needs to be addressed programmatically. Digital Preservation is more than a back-up.

**Why do we need it?**

Digital files become inaccessible over time unless they are managed and migrated to new technology as needed. Even content that resides on a web site or social media site is at risk.
Why Should Digital Preservation Matter to You?

Everyone has a role to play in safeguarding our university’s digital assets.

The intellectual property of our institution is at risk. We must have an active digital preservation program in place to be a leading institution. Without such a program all the content institution members create is at risk of being lost.

The landscape for how institutional memory and culture are created has changed dramatically. Current estimates are that without digital preservation electronic records could become inaccessible within a few years, a decade at most. To ensure long-term viability and relevance of our work we need to begin a digital preservation program now as a core component of campus activities like teaching or research.

Facts about Digital Preservation

- Faculty, students and staff are creating valuable content daily (e.g., photo and other media archives, works of artistic and scholarly expression). Without a program in place to ensure the viability of that content over the long-term, those files will become unreadable.

- Federal granting agencies require data management plans for grants. Digital preservation is a core element of a data management plan.

- We compete with institutions where digital preservation is ongoing, potentially putting us at a disadvantage when seeking funding. Implementing a digital preservation program will give us a competitive advantage over institutions where projects and programs are not yet in place.

- Digital preservation isn’t just about judicious backing up of files. We need to preserve these works in order to connect our alumni to their campus experiences and to show other benefactors evidence of our community’s work.

What is Digital Preservation?

Digital preservation combines policies, strategies and actions to ensure long-term access to content that is born digital or converted to digital form regardless of the challenges of file corruption, media failure and technological change.

Digital preservation is an ongoing process, not a one-time activity, and needs to be addressed programmatically. Digital Preservation is more than a back-up.

Why do we need it?

Digital files become inaccessible over time unless they are managed and migrated to new technology as needed. Even content that resides on a web site or social media site is at risk.
Intended for: Business Administrator

Why Should Digital Preservation Matter to You?

Everyone has a role to play in safeguarding our university’s digital assets.

The intellectual property of our institution is at risk. We must have an active digital preservation program in place to be a leading institution. Without such a program all the content institution members create is at risk of being lost.

Current estimates are that without digital preservation electronic records could become inaccessible within a few years, a decade at most. To ensure long-term viability and relevance of our work and to mitigate legal and financial risk we need to begin a digital preservation program now.

Facts about Digital Preservation

➤ Digital preservation will maintain auditable records and secure the university’s historical record.

➤ New and pending legislation at state and federal levels increasingly obligate our institution to investigate digital preservation, so acting now will put us ahead of the requirements.

➤ Federal granting agencies require data management plans for grants. Digital preservation is a core element of a data management plan.

➤ We compete with institutions where digital preservation is ongoing, potentially putting us at a disadvantage when recruiting faculty, students and staff. Implementing a digital preservation program will give us a competitive advantage over institutions where projects and programs are not yet in place.

➤ In the event of a FOIA request a comprehensive digital preservation program will ensure we have discoverable records preserved for the long-term.

➤ Digital preservation can be implemented incrementally, with greater benefits achieved through higher levels of implementation.

What is Digital Preservation?

Digital preservation combines policies, strategies and actions to ensure long-term access to content that is born digital or converted to digital form regardless of the challenges of file corruption, media failure and technological change.

Digital preservation is an ongoing process, not a one-time activity, and needs to be addressed programmatically. Digital Preservation is more than a back-up.

Why do we need it?

Digital files become inaccessible over time unless they are managed and migrated to new technology as needed. Even content that resides on a web site or social media site is at risk.

12/9/2013
Why Should Digital Preservation Matter to You?

Everyone has a role to play in safeguarding our university’s digital assets.

The intellectual property of our institution is at risk. We must have an active digital preservation program in place to be a leading institution. Without such a program all the content we create is at risk of being lost.

Current estimates are that without digital preservation electronic records could become inaccessible within a few years, a decade at most. To ensure long-term viability and relevance of our work we need to begin a digital preservation program now as a core component of campus activities like teaching or research.

Facts about Digital Preservation

- Twenty-first century research requires management of data, mining of data and working with libraries and IT staff to manage data so it’s available long-term.

- Technology changes. Without a program for long-term preservation the intellectual content you are developing today will become inaccessible to the researchers of tomorrow.

- Content creators have the most to lose without a preservation program in place. Campus administrators at all levels need to understand the value you place on digital preservation and allocate resources accordingly.

- Digital preservation isn’t just about judicious backing up of files. Initially you will need to invest time in selecting and organizing files for archiving, but after the initial push the ongoing time commitment

What is Digital Preservation?
Digital preservation combines policies, strategies and actions to ensure long-term access to content that is born digital or converted to digital form regardless of the challenges of file corruption, media failure and technological change.

Digital preservation is an ongoing process, not a one-time activity, and needs to be addressed programmatically. Digital Preservation is more than a back-up.

Why do we need it?
Digital files become inaccessible over time unless they are managed and migrated to new technology as needed. Even content that resides on a web site or social media site is at risk.
Intended for: Information Management (Librarian)

Why Should Digital Preservation Matter to You?

Everyone has a role to play in safeguarding our university's digital assets.

The intellectual property of our institution is at risk. We must have an active digital preservation program in place to be a leading institution. Without such a program all the content we create is at risk of being lost.

Current estimates are that without digital preservation electronic records could become inaccessible within a few years, a decade at most. To ensure long-term viability and relevance of our work we need to begin a digital preservation program now as a core component of campus activities.

What is Digital Preservation?

Digital preservation combines policies, strategies and actions to ensure long-term access to content that is born digital or converted to digital form regardless of the challenges of file corruption, media failure and technological change.

Digital preservation is an ongoing process, not a one-time activity, and needs to be addressed programmatically. Digital Preservation is more than a back-up.

Why do we need it?

Digital files become inaccessible over time unless they are managed and migrated to new technology as needed. Even content that resides on a web site or social media site is at risk.

Facts about Digital Preservation

➤ As members of the institution realize the importance of digital preservation, the job of organizing and managing information changes. The formats have changed but the role of the information management professional is the same.

➤ You are in a trusted position and can help move digital preservation forward through collaboration and communication.

➤ Faculty and others might not understand the difference among backup, storage and preservation or how the idea of an Institutional Repository differs from digital preservation.

➤ We know our institution is at risk without a digital preservation program. We need our colleagues to understand the issues and the role they can play in managing and purging files to keep only what's important.

➤ The library has a crucial role to play in managing the files under your purview. Be clear about your role and the role of others in the
Why Should Digital Preservation Matter to You?

Everyone has a role to play in safeguarding our university's digital assets.

The intellectual property of our institution is at risk. While we ensure that our network is secure and we have reliable backup and virus protection, we do not have a digital preservation program to safeguard our intellectual property. This puts all the content we create at risk of being lost.

Current estimates are that without digital preservation electronic records could become inaccessible within a few years, a decade at most. To ensure long-term viability and relevance of our work we need to begin a digital preservation program now.

Facts about Digital Preservation

➢ Digital preservation is a core service that needs to be managed in concert with other IT programs.

➢ The IT Department already plays a key role in ensuring content viability campus-wide, tracking and notifying staff of software changes so they can make adjustments. An extension of that role is to monitor hardware obsolescence and ensure files are migrated to new formats—a critical component of a digital preservation program.

➢ University policies and procedures for managing data need to be expanded to include digital preservation guidelines. Guidelines can help clarify the appropriate selection of files, thereby reducing the amount of storage space needed.

➢ Digital preservation ensures that files will continue to be accessible over time, not that they are just backed up and stored.

➢ The university needs to identify multiple locations for storing preserved files. The archive should not be maintained solely on our institution’s servers.

➢ Digital preservation is programmatic and involves ongoing costs. Plans need to include server maintenance and replacement.

What is Digital Preservation?

Digital preservation combines policies, strategies and actions to ensure long-term access to content that is born digital or converted to digital form regardless of the challenges of file corruption, media failure and technological change.

Digital preservation is an ongoing process, not a one-time activity, and needs to be addressed programmatically. Digital Preservation is more than a back-up.

Why do we need it?

Digital files become inaccessible over time unless they are managed and migrated to new technology as needed. Even content that resides on a web site or social media site is at risk.
Internet Archive is a way to archive public domain materials free of charge. It is important to have multiple backups of digital files in case of unexpected loss of originals.

Table of Contents

Navigating to the Website ..................................................2
Creating an Account .........................................................3
Logging into Internet Archive ..............................................6
Uploading a Digital Object ..................................................7
Uploading a Collection of Objects .........................................11
Viewing Previously Uploaded Objects .................................14
Downloading Previously Uploaded Objects .........................16
Navigating to the Website

Open your browser (Internet Explorer, Google Chrome, or Mozilla Firefox are all options)

Click on the URL bar

Type www.archive.org

Then hit the enter key
Creating an Account

It is best to create a user account for Internet Archive in order to make the best of all offered features before uploading any digital objects to the website.

1. Locate the "join us" link near the top right of the Internet Archive page.

2. Click on the "join us" link and fill in the required information.

Get a Virtual Library Card

Complete the form below to create a new Internet Archive account:

Your email address: 

Choose a screen name: 

Choose a password:

Confirm password:

Internet Archive announcements:

☐ Send me general announcements from the Internet Archive (approx. one per month)
☐ Send me announcements about Internet Archive events in the San Francisco Bay Area

Terms of Use:

☐ I’ve read and agreed to Internet Archive’s Terms of Use

[Get Library Card]
Using Internet Archive:
A guide created by the Digital POWRR Project

a. Add your e-mail address. It must be an e-mail address that you have access to check so you can confirm the address.

Your email address verification required
digitalpowrr@gmail.com

b. Choose a screen name. This will appear if you decide to write reviews or have other interactions publicly on Internet Archive.

Choose a screen name displayed with your reviews, etc.
DigitalPOWRR

c. Create a strong password, including a symbol such as !@#$%^ will greatly strengthen your password.

Choose a password

**********

d. Re-type the same password from step C into the confirm password field.

Confirm password

**********

e. Decide if you would like to receive e-mails from Internet Archive, if so select the boxes. If not, feel free to leave them blank.

Internet Archive announcements:

- Send me general announcements from the Internet Archive (approx. one per month)
- Send me announcements about Internet Archive events in the San Francisco Bay Area

f. Read the terms of use and check the agreement box.

Terms of Use:

- I've read and agreed to Internet Archive's Terms of Use

g. Select the "Get Library Card" button.
3. A confirmation screen should appear in your browser.

Verification Email Sent

We've sent an email to digitalpowrr@gmail.com. You'll need to read that and click on the verification link to finish creating your account.

Please check your Spam folder if you do not see the verification email in your inbox.

4. Log in to your e-mail account to read the verification e-mail.

5. Click the link given in the e-mail.
Using Internet Archive:
A guide created by the Digital POWRR Project

6. After a few moments you should receive “Verification Successful” screen.

Logging into Internet Archive
If you have not yet created an account please follow those steps first.
1. Navigate to the Internet Archive Homepage
2. Click on the “login” link located near the top right corner of the page.
3. Enter the e-mail address and password you choose when registered.
4. Click “Log In”.
5. If you are successful it will show your chosen username near the top right corner.

Created by: Digital POWRR Project
For more information visit digitalpowrr.niu.edu
Using Internet Archive:
A guide created by the Digital POWRR Project

**Uploading a digital object**

1. Click on the “Upload” button located on near the top right corner of the Internet Archive homepage. Make sure you are logged in before doing so.
Using Internet Archive:
A guide created by the Digital POWRR Project

2. A “Share your Files” screen should appear. Click on the “Upload Files” button.

3. You may drag and drop the file from an open file explorer window, or you can click on the “Choose Files to Upload” button.

4. You may also drag the file from the “Choose files to upload” menu and drop them in the grey area. Once the file has been selected you may either click “Cancel” if you have already dragged and dropped the object, or “Open” if

Created by: Digital POWRR Project
For more information visit digitalpowrr.niu.edu
Using Internet Archive:
A guide created by the Digital POWRR Project

you prefer not to drag and drop.

5. For the digital object you will want to edit the metadata available. For each area you can click on the area and edit the information as you wish. When you feel that the information is sufficient you can click the "Upload and

Created by: Digital POWRR Project
For more information visit digitalpowrr.niu.edu
Using Internet Archive:
A guide created by the Digital POWRR Project

Create Your Item” button.

6. You may receive a few different loading screens; just give Internet Archive a few moments to work.
Using Internet Archive:
A guide created by the Digital POWRR Project

7. You will then reach a confirmation page that the page has successful been created.

8. If you notice a mistake you may edit the item by clicking the “Edit Item” link.
Using Internet Archive:
A guide created by the Digital POWRR Project

Uploading a Collection of Objects

1. Click on the "Upload" button located on near the top right corner of the Internet Archive homepage. Make sure you are logged in before doing so.

![Upload button](Image)

Universal access to all knowledge

Hello DigitalPOWRR (not you? sign in or log out)

Welcome to the Archive

The Internet Archive, a 501(c)(3) non-profit, is

2. A "Share your Files" screen should appear. Click on the "Upload Files" button.

Share your Files
Possible contributions: audio, video files that you have the rights to share.

Or choose an alternate upload method:

- Classic Upload
- Upload Form

Created by: Digital POWRR Project
For more information visit digitalpowrr.niu.edu
3. You may drag and drop the files from an open file explorer window, or you can click on the “Choose Files to Upload” button.

4. You may also drag the files from the “Choose files to upload” menu and drop them in the grey area. Once the file has been selected you may either click “Cancel” if you have already dragged and dropped the object, or “Open” if you prefer not to drag and drop. Make sure that all necessary items for the collection are listed in the grey area.
5. For the digital objects you will want to edit the metadata available so that it reflects the entire collection. For each area you can click on the area and edit the information as you wish. When you feel that the information is sufficient you can click the "Upload and Create Your Item" button.

6. You may receive a few different loading screens; just give Internet Archive a few moments to work.
Using Internet Archive:
A guide created by the Digital POWRR Project

7. You will then reach a confirmation page that the page has successful been created.

8. If you notice a mistake you may edit the item by clicking the “Edit Item” link.
Using Internet Archive:
A guide created by the Digital POWRR Project

Viewing Previously Uploaded Objects

1. Make sure you are logged in to Internet Archive.

2. Click on the “Account” link located on the black navigation bar.

3. Under the “Your Activity” area, click on the “Uploads” link.

4. Your previous uploads will be listed. If an item or collection that you’ve recently added is missing it may not have finished going through the official process yet. Wait a couple of hours and check back again later.
Download Previously Uploaded Objects

1. Follow the steps for "Viewing Previously Uploaded Objects" (see page 15)
2. From "Search Results" choose the object/collections that you wish to download/recover and click on the title you gave the object/collection.

3. Depending on the format of the objects your screen may look a little different. For example texts versus pictures files.
4. Locate the format of the object you would like to download/recover on the right hand side of the screen, and click on that link.

5. Depending on the browser you are using the download will appear in a different location. If you do not see the download in your browser check the Downloads folder on your computer. The following image is from downloading using Google Chrome.

6. You may download it in more than one format if you wish. Images may not automatically download; instead they may open in a new browser window. If so you can right click on the image and select “Save image as...” and save the
Using Internet Archive:
A guide created by the Digital POWRR Project

object in the location of your choice.

Created by: Digital POWRR Project
For more information visit digitalpowrr.niu.edu
Intellectual Capital at Risk: Data Management Practices and Data Loss by Faculty Members at Five American Universities

Jaime Schumacher
Northern Illinois University

Drew VandeCreek
Northern Illinois University

Abstract

A study of fifty-six professors at five American universities found that a majority had little understanding of principles, well-known in the field of data curation, informing the ongoing preservation of digital materials and chose to manage and store work-related data by relying on the use of their own storage devices and cloud accounts. It also found that a majority of them had experienced the loss of at least one work-related digital object that they considered to be important in the course of their professional career. Despite such a rate of loss, a majority of respondents expressed at least a moderate level of confidence that they would be able to make use of their digital objects in twenty-five years. The data suggest a strong correlation between faculty members' digital-object loss and their data management practices. University professors producing digital objects should be aware of fact that these materials are subject to loss. They should also know that the implementation of better data management practices and the development of programmatic digital curation efforts on university campuses show great promise for mitigating university professors' loss of important work-related digital materials.

Received dd Month yyyy

Correspondence should be addressed to Drew E. VandeCreek, Founders Library, Northern Illinois University, DeKalb, IL 60115 USA. Email: drew@niu.edu.

The International Journal of Digital Curation is an international journal committed to scholarly excellence and dedicated to the advancement of digital curation across a wide range of sectors. The IJDC is published by the University of Edinburgh on behalf of the Digital Curation Centre. ISSN: 1746-8256. URL: http://www.ijdc.net/

Copyright rests with the authors. This work is released under a Creative Commons Attribution (UK) Licence, version 2.0. For details please see http://creativecommons.org/licenses/by/2.0/uk/

International Journal of Digital Curation
2014, Vol. 0, Iss. 0, 1–

http://dx.doi.org/10.2218/ijdc.v0i0.0
DOI: 10.2218/ijdc.v0i0.0
Introduction

In the past twenty years college and university faculty members have made digital materials integral parts of their work. A scholar's collection of work-related digital content often includes important, unique research data or other products of the creative process, as well as materials important to her/his teaching activities and committee work. As digital objects have become increasingly ubiquitous, many librarians and other information professionals have come to warn that they are very prone to loss. Acknowledging this reality, federal granting agencies in the United States have recently begun requiring scholars seeking financial support to describe in their proposals how they plan to provide for the long-term preservation and management of data sets and other materials produced in the course of research activities. (National Science Foundation [NSF], 2012). In the interest of maintaining their faculty members' eligibility for grant support, many colleges and universities have developed digital data management and preservation measures meeting or exceeding granting agencies' specifications. They will secure an important portion of college or university scholars' intellectual capital, but what of their faculty members' other work-related digital objects? Within this mass of materials there surely exists other digital content of great significance to individual scholars and their institutions. This paper presents the findings of an investigation into the state of digital object management and preservation among fifty-six professors at five universities. It considers their responses, in an interview format, to a set of several questions: "What types of work-related materials do faculty members possess, and in what formats?" "What materials would they most like to recover in the event of an apparent loss?" "How do they presently manage and store their work-related digital materials?" "What are their expectations about these materials'
future availability?” and, finally, “What rates of digital object loss have they experienced?”

Our study found that participating faculty members produced a wide array of digital objects in the course of their professional work. As a group, they identified research data and other products of their scholarly activities, teaching resources, and administrative materials as those they would most like to recover in the event of an apparent loss. A large majority of participating faculty members reported that they stored and managed their work-related digital materials themselves, relying on a collection of free-standing devices, such as personal computers and portable hard drives, as well as cloud accounts, rather than their institution’s available network capacity. Many of those not using network storage reported that they were unaware of its existence or availability, or found its use inconvenient in some respect. In this context, a majority of respondents reported that they had experienced the loss of at least one work-related digital object that they considered to be important in the course of their academic career. Upon interviewers’ review of the leading causes of digital object loss, a large majority of participants reported that they believed that their data loss was actually greater than they had initially indicated. Of those reporting both kinds of loss, a disproportionate number were among the individuals not using network storage. Despite such a rate of loss, a majority of respondents expressed at least a moderate level of confidence that they would be able to make use of their digital objects well into the future.

These results indicate that digital-object loss is a common, if not regular, component of academic life. They show a correlation between faculty members’ use of network storage and the survival of their digital objects. Conversely, they also reveal a link between a sole reliance on non-networked devices and services for digital-object storage and the more frequent loss of digital objects. Many participants’ failure to make use of
available network storage, as well as upward revisions of digital-object loss estimates after a review of the major potential causes of loss, show that many faculty members have little understanding of the factors placing digital materials at risk of loss or how they might mitigate that risk. They also demonstrate a strong connection between this lack of knowledge of risk factors and basic data management practices and digital-object loss itself. Finally, participants’ reported expectations about the future availability of their digital objects suggest that university professors are often very over-optimistic in this regard, a phenomenon that shows a correlation with their lack of knowledge of the causes of digital object loss.

These results suggest that many individual faculty members can take a first step toward more effective digital-object preservation by using their institution’s available network capacity, which is, in the vast majority of cases, backed-up in a regular and systematic manner. Colleges and universities can encourage network usage by increasing their efforts to bring its availability to faculty members’ attention and reducing barriers to its convenient use. Our study’s broader portrait of the precarious state of faculty-produced digital data on five campuses can also perhaps help those advocating their institution’s more effective, programmatic preservation of selected faculty-produced, non-grant related digital objects to persuade colleagues and administrators to provide the financial support needed to underwrite the necessary activities. Finally, however, our results suggest that many faculty members’ data management practices may present a significant obstacle to attempts to curate and preserve their digital objects in a more effective, programmatic manner. Any attempt to integrate faculty-produced digital objects stored outside of a college or university’s network into an institution-wide digital preservation program will require participating faculty members to devote a great deal of time and effort to the location, retrieval, and
submission of their materials. It will also oblige information professionals taking part in the planning and administration of a digital preservation program to handle larger amounts of materials, usually not identified in any standard manner, and often lacking important metadata, than they would otherwise encounter. In this light, digital preservation planners must carefully weigh the pros and cons of any attempt to integrate faculty-created digital objects residing outside an institution’s network into a digital preservation initiative. In cases where the time and expense required prohibit such an attempt, these materials will at best reside in a sort of limbo, neither demonstrably lost nor preserved in any programmatic, effective manner.

Literature Review

The literature exploring digital object loss and potential solutions to it has largely framed the problem in general terms, emphasizing that such materials created in past years are often not compatible with today's hardware or software. In addition, it reminds us that storage media is subject to failure, particularly in cases of neglect or damage, in which the basic binary constructs that make up digital materials can lose their integrity (Pogue, 2009). In exploring how this risk affects institutions of higher education, researchers have seldom provided empirical studies of rates of actual data loss in this population. In 2008 one investigator in the field called for more work devoted to “quantifying the extent of digital information loss or compromise, or, at the very least, to document more examples to supplement the few specific studies currently available.” (Harvey, 2008, p. 2). Six years later, digital preservation professionals can still benefit from this type of empirical study.

Work providing information about the actual rate of loss of scholarly digital materials in a university setting is hard to find. In 2006 the Digital Preservation
Coalition (UK) presented *Mind the Gap: Assessing Digital Preservation Needs in the UK*, a report based on a survey of "a wide range of organisations in different sectors," including education, libraries, archives, museums, local and central government bodies, scientific research institutions, as well as organizations in the pharmaceutical, financial, manufacturing, engineering, media, energy and chemical, and publishing industries (Waller & Sharpe, 2006, p. 11). In it, researchers reported that only 29% of respondents to a 2005 survey reported that they had "not lost access to some digital information as a result of it being impossible or too expensive to recover. Even when referring to their most important type of data, this proportion only rose to 43%." (Waller & Sharpe, 2006, p. 18). In a 2013 study, a team of Canadian scientists working at universities and research institutes examined the availability of data sets, ranging from two to over twenty years old, finding that amidst prevailing data management practices "the odds of a data set being extant fell by 17% per year," and concluding that "the availability of research data declines rapidly with article age." (Vines et al., 2013, 19).

The available literature provides a clearer picture of university professors' current data management practices and understanding of the life cycle of digital objects. A 2012 study found that 95% of faculty members at a medium-sized university reported that they believed that they should be personally responsible for their data (Scaramozzino, J. M., Ramirez, M. L., & McGaughey, K. J., 2012, 356). A 2013 study of faculty members at American universities and colleges conducted by Ithaka (the organization behind JSTOR and Portico) found that 80% of respondents in the Sciences, and nearly 80% in the Humanities and Social Sciences, preserved research data themselves, using commercially or freely available software or services (Housewright, Schonfeld, & Wulfson, 2013, 63). The University of North Carolina's 2012 report *Research Data Stewardship at UNC: Recommendations for Scholarly Practice and Leadership* found
that "While some (faculty members) save data in repositories or centralized servers, others relied on external hard drives or CDs for backup. Beyond the research project period, some faculty admitted that few, if any steps were taken to preserve their data long-term." (School of Information and Library Science at University of North Carolina [SILS UNC], 2012, 10). A 2011 publication concluded that that a group of academic archaeologists and arts historians generally preserved their images on such "an ad hoc basis," and concluded that they "generally did not understand preservation issues surrounding their images...." (Beaudoin, 2011, 493). Two 2009 reports also concluded that most scholars lacked the knowledge of effective practices necessary to manage their digital objects in an effective manner. (Martinez-Uribe; RIN & the British Library). Nonetheless, one recent study has found that over one-third of researchers at a large American university expected their digital objects to remain usable after "ten years or more." (Fear, 2011, p. 62).

This study explores how faculty members’ data management attitudes and practices may affect future attempts to integrate some of their digital objects into campus-wide digital preservation programs. Beaudoin has observed that attempts to integrate faculty members’ collections of work-related images into an institutional repository and preservation program “could complicate the work of librarians,” and goes on to discuss intellectual property issues and a likely dearth of metadata as major causes of these difficulties. (Beaudoin, 2011, 493). The present study turns from these issues to assess what impact faculty members’ reliance on non-networked storage devices may have on attempts to integrate their data into a digital preservation program. Many standards and best practices published in the field of digital preservation call for institutions to manage selected materials as a whole in “an archive, repository, data centre, or other custodian” or simply assume that this has been done. (Digital Curation Centre [DCC], n.d.) To
date, no studies have explored what implications college and university faculty
members’ data management attitudes and practices might have for the seemingly
straightforward work of bringing together a set of digital objects for programmatic
digital preservation measures.

Finally, this article examines the relationship between college and university faculty
members’ data management attitudes and practices and their loss of digital objects.
More specifically, it asks if the use of network storage can provide higher levels of
preservation for faculty-produced digital objects and explores factors bearing on faculty
members’ use of it. In 1996 and 1997, three scholars studied factors influencing
engineering professors’ use of computer networks for e-mail, electronic discussion
groups, accessing remote databases, accessing remote computer facilities and file
transfer. (Abels, Liescher & Denman, 1996; Liescher, Abels & Denman, 1997). Their
work did not examine faculty members’ use of network storage. The authors know of no
scholarly inquiries into the relationship between faculty members’ use of network
storage and the preservation of their work-related digital objects, however.

Research Methodology

In 2012 and 2013 members of a team studying how medium-sized and smaller
institutions of higher education might begin to address the issue of digital preservation
explored faculty members’ creation and use of digital objects, as well as their
understanding of preservation issues pertaining to them. This study was conducted with
the support of the Institute of Museum and Library Services (National Leadership Grant
05-11-0156-11), at five Illinois universities: Northern Illinois University, Chicago State
University, Illinois State University, Illinois Wesleyan University, and Western Illinois
University. Potential faculty participants included grant applicants and recipients (as
provided by the institutions' Offices of Sponsored Projects), as well as scholars whose publications resided in an institutional repository. Drawing on a pool of 119 candidates contacted via an email message requesting a 30 minute, in-person interview, team members conducted 56 faculty interviews, the results of which were de-identified. Participating individuals represented a wide range of disciplines, including specialists in the humanities, the physical sciences, the biological sciences, the social sciences, engineering and education. During the individual interviews, participants were asked to describe their professional activities, the nature of the digital information they had created in the course of their work and the types of data formats included in their body of work. They were then asked to describe how they stored and managed their data, and to identify those materials they considered to be most valuable and hence would most want to recover in the event of their apparent loss. Subsequent questions asked participants to report experiences of data loss (if any), then asked them to review their report in light of a discussion of leading causes of digital-object loss. Finally, interviewers asked participants to describe their level of confidence in the long-term availability of their digital files.

Findings

Study participants provided the following data. (VandeCreek and Schumacher, 2014). When asked to identify up to three sets of electronic files they would attempt to recover first in the event of an apparent loss, the majority of participants selected materials pertaining to four major functions of university life. Forty-two participants (75%) identified electronic files within the category of scholarly materials (including research data and scholarship in a broader sense); twenty-three (41%) designated teaching materials; ten (17.8%) selected administrative and/or organizational materials; and five
(8.9%) named electronic communications.

![Categories of Digital Materials Rated as a High Priority for Protection Against Loss](image)

**Figure 1**

Having identified up to three categories of digital materials that they would most want to recover in the event of an apparent loss, participants were then asked to list any and all file types that made up the objects within each category. They named JPEG, PDF and .doc files most often, but answers reporting the use of a file type not found in the study’s roster of twenty-eight common types ("other") were fourth most common. Within this group, several participants mentioned both proprietary and custom formats as well as types unique to software programs designed for use in their specific area of expertise. A full set of participant responses is shown in Figure 2.
Faculty members’ responses to inquiries about how they managed and stored such highly-valued, work-related digital objects revealed that most managed their data themselves, relying on some combination of individual devices and storage services. Again, participants were asked to name any and all digital object storage and preservation methods that they used. Thirty-seven (66%) relied on the hard drives of their office computer; twenty-two (39%) used an external hard drive; twenty-one (37.5%) used a hard drive as a built-in component of a personal computer; eighteen (32%) used cloud-based services; sixteen (28.5%) used a Flash/USB drive; ten (17.8%) used their email account(s); six (10.7%) used means or devices not mentioned in the project interview’s list of storage options; and three (5.4%) relied on optical discs like CDs or DVDs. Few faculty members interviewed took advantage of opportunities to back up their materials in more secure environments. Twenty (35.7%) employed institutional network capacity made available for the storage of work-related materials.
One (1.8%) participant made use of a discipline-specific external repository, meaning that thirty-five (62.5%) relied entirely on free-standing, non-networked devices for digital object storage.

![Storage/Back-up Methods Used](image)

Figure 3

Participating faculty members’ data management practices in many cases resulted in a loss of digital materials that they considered to be important to their professional activities. Thirty-one (55.3%) of participants indicated that they were aware that they had lost such work-related digital objects and been unable to replace them with backup files in the course of their professional career.
Participants Reporting Significant Data Loss

![Bar chart showing participants reporting significant data loss.](image)

Figure 4

Researchers then asked participants if they possessed digital content that, despite having made no recent attempts to access the materials, they would likely be unable to open and use in light of researchers' descriptions of common causes of digital data loss other than the simple failure of storage media. These causes, as noted above, could include files created within older software or applications to which the creator no longer had access or materials stranded on media that was incompatible with current hardware, like floppy discs, zip discs, or obsolete hard drives. Thirty-five (62.5%) participants reported that they did indeed have digital materials that were likely lost to use (despite still having the materials at hand), a situation researchers referred to as a "previously unrealized data loss."
Participants Reporting Previously Unrealized Data Loss*

*Despite having the data at hand, participants would be unable to open and use digital files that were created with out-of-date software or stranded on obsolete media, like floppy disks.

Those participants eschewing the use of university-furnished network storage were particularly susceptible to data loss. Of the thirty-one reporting knowledge of a past data loss, twenty-three (74%) were among those relying entirely on free-standing devices, optical discs and external services, usually in some combination. Of the thirty-five individuals reporting an unrealized data loss, nineteen (54.2%) made it clear that they were relying only on non-networked devices and accounts for their data storage.

Faculty members offering comments in addition to their responses to the set of standard interview questions reported that a number of factors contributed to their limited use of university-provided network capacity and heavy reliance on stand-alone storage devices. One participant stated that s/he had never been informed of the existence of network storage capacity allocated to her/his department. Another stated that s/he believed that the university “should have mass storage for departments,”
indicating a lack of awareness of its actual availability at her/his university. A third stated that although s/he was aware of the existence of university-provided network storage, s/he had not been offered the use of it. Two members of a department took issue with institutional policies governing the use of their department’s network, one stating that it was “too locked down.” The other indicated that her/his unwillingness to work with a network administrator that s/he found unhelpful also contributed to her/his decision to store and manage her/his own data using a number of free-standing devices. A scholar producing digital objects each nearly a terabyte in size noted that the university’s available network storage failed to provide adequate space for her/his needs. Another stated that s/he believed that her/his institution’s network infrastructure was unable to move files of the size that s/he regularly used in a timely fashion. Finally, one professor indicated that s/he was aware of university-provided network storage space available to her/him, but did not make use of it because s/he could not access it from home.

The last remark led project researchers to investigate the terms on which participating institutions made secure network storage available to their faculty members. Inquiries discovered that faculty members at Illinois State University had ready access to network facilities available from remote locations via a Web VPN authentication system at the time of the project survey. (M. Walbert, personal communication, September 26, 2013). At Western Illinois University faculty members also enjoyed off-campus access to network storage space, via an SFTP client, at the time that researchers gathered data for this study. (J. Hancks, personal communication, September 30, 2013). At Northern Illinois University, Chicago State University and Illinois Wesleyan University, faculty members enjoyed access to secure network storage space, but could not access it from remote locations at the time of the study. (T. Lusher,
personal communication, October 17, 2013; M. Kong, personal communication, October 8, 2013; T. Short, personal communication, October 22, 2013).

Faculty members able to use college or university-administered networks from off-campus reported a higher level of network usage. At the two institutions providing professors with off-campus access to materials stored on a university network, seven of fourteen (50\%) respondents reported that they saved digital materials on that network. At the universities that did not offer remote access to materials stored on their networks, only thirteen of forty-two (31\%) made use of available network storage capacity.

Faculty members able to use college or university-administered networks from off-campus also reported slightly lower levels of data loss. At institutions providing off-campus access to available network storage capacity, ten of the fourteen (71\%) individuals surveyed reported a data loss of some kind. At institutions without remotes network access available, thirty-four of the forty-two (81\%) individuals surveyed stated that they had lost data.

In this context, participating faculty members expressed considerable confidence that they would be able to recover (retrieve and use) work-related digital objects in the event that they discovered their apparent loss at the time of the interview. Of a total of 56 individuals interviewed, eleven (19.6\%) indicated that they were “very confident” that they would successfully recover their digital objects; eighteen (32.1\%) reported that they were “mostly confident;” and seventeen (30.4\%) declared that they were “somewhat confident.” Four (7.1\%) faculty members stated that they were “slightly confident;” and five (8.9\%) expressed the fact that they were “not at all confident” in their ability to recover their digital objects. One replied that s/he did not venture to estimate her/his level of confidence in the future retrieval of digital objects. Thus fully eighty-five percent of participating faculty members expressed at least a modest
("somewhat" or better) level of confidence in their ability to retrieve and use digital objects existing at the time of the survey, with 51.8% of them indicating that they were very or mostly confident in this outcome.

How Confident are you that you can Recover your Data?

![Bar Chart]

Figure 6

In response to an inquiry asking if they were confident that they would be able to make use of their present digital objects in twenty-five years, six (10.7%) participating faculty members reported that they were “very confident,” twelve (21.4%) concluded that they were “mostly confident,” and sixteen (28.6%) were “somewhat confident.” Eight (14.2%) reported that they were “slightly confident” and twelve (21.4%) stated that they were “not at all confident” that they would be able to recover their digital objects at that time. Two indicated that they did not know if they would be able to recover their digital objects in twenty-five years. In this matter, a total of 32.1% reported that they were “very” or “mostly” confident to recover their data twenty-five years in the future, 60.7% expressed that they were at least “somewhat” confident in the
same outcome, and fully 75% expressed any confidence in that prospect.

**How Confident Are You That You Can Recover Your Data in 25 Years?**

![Bar chart showing confidence levels](chart.png)

Figure 7

**Discussion**

The data presented in this study shed light on several topics of interest to information professionals studying digital preservation or simply seeking to provide enhanced levels of preservation for collected materials in a college or university environment. Many participants’ descriptions of using an *ad hoc* system of free-standing computers, hard drives and cloud accounts corroborated earlier studies showing that other groups of university professors had adopted very similar personal data management practices. (Housewright, et al; SILS UNC).

Study participants’ widespread disinclination to make use of available network storage and their frequent acknowledgment that they did not understand the major causes of digital object loss supported the findings of previous work (Beaudoin;
Martinez-Uribe; RIN & the British Library) showing that other groups of university professors struggled to grasp the causes and risk of digital-object loss and lacked the knowledge of data-management practices necessary to increase the likelihood that digital objects would be available for future use.

In addition, this study provided empirical information about actual data loss, as requested by Ross Harvey. Its survey of university professors demonstrated that a majority of participants lost the use of work-related digital objects, a result supporting the findings of the Digital Preservation Coalition’s Mind the Gap study and Canadian scientists’ review of the availability of research data over time (Vines, et al.). This study’s results suggested that many participants’ lack of awareness of their data’s susceptibility to loss and free-wheeling data management practices may have exacerbated the risks associated with software and hardware incompatibility and storage media failure. Responses showed that those storing and managing digital objects in a collection of free-standing devices and accounts alone showed a larger rate of data loss than those who made use of their institution’s network. The authors of this paper believe this to be the first instance in which a study has explored possible relationships between faculty members’ understanding of data management issues, their data management practices and their reported rates of actual data loss.

When considering how, if at all, participants’ ability to make use of network storage from remote locations influenced the rate at which they used such institution-provided capacity, this study found that faculty members enjoying off-campus network access at their institution had a moderately higher rate of network use. Likewise, those participants taking advantage of the opportunity to use their network from off campus experienced a slightly lower rate of data loss than those who ignored it. Again, the authors of this paper believe that this is the first research of its kind to be discussed in a
publication.

The study's data also provided evidence supporting and augmenting a recent finding that many university professors at a single institution believed that they would be able to recover and use seemingly-lost digital objects today and in the future (Fear). In that research the author found that more-confident participants expected to be able to recover and use their digital materials after a period of ten years. The present study showed that a sizable number of participating faculty members expected their digital materials to be intact and usable in twenty-five years. The great gap between basic best practices developed by information science professionals and many participating professors' actual data management practices suggests that faculty members will continue to lose the use of digital materials in the future. In this light, many participants' assumption that their digital objects will remain available for use at such a distant date indicates that librarians, archivists and other professionals aware of digital objects' susceptibility to loss should take it upon themselves to initiate or increase efforts to educate their colleagues of the dangers they face.

This study's findings suggest that even as colleges and universities move to create digital preservation policies and measures necessary to meet funding agencies' data management mandates, a need for programmatic digital preservation measures also exists at an even broader institutional level: that pertaining to faculty members' digital materials not bound by a granting body's preservation mandate. Information professionals now enjoy a remarkable opportunity to enhance the level of service they and their larger libraries and archives provide to their university or college. They already possess the foundational knowledge on which an institution's digital preservation program can be established and developed, and can gain additional expertise and skills within the established channels of their professional discourses.
(Fyffe, Ludwig, & Warner pp. 11-12).

This study’s findings reveal a major challenge that confronts professionals seeking to build a digital preservation program on such an expanded, institutional level, however. Discussions of best practices for digital object preservation direct those beginning a new program to bring available digital objects together in a single repository and compile an inventory of them. (DCC; NDSA; Angevaare, 2009). The collection of materials in a single repository facilitates their management and evaluation. Professors using a college or university network for digital-object storage can readily submit materials for evaluation and potential preservation measures simply by moving an object and its metadata from one location on the network to another made available to them. Any attempt to include other faculty-created materials in an institution-wide digital preservation program will need to rely on professors’ willingness to search their free-standing devices and cloud accounts for digital objects potentially appropriate for preservation and transfer the materials to digital preservation planners, however. These activities may often be difficult and unpleasant for a faculty member, requiring them to step away from work schedules already packed with teaching, research, student advising and committee service, and perhaps sacrifice time devoted to non-professional activities. The search for materials will require an examination of media and devices that many have already conceded are disorganized and ineffective, and the discovery that they have lost data may make a professor feel foolish. In many cases the search will succeed only as faculty members are able to locate devices capable of reading obsolete media types and equipped with software capable of reading obsolete or unsupported applications. Digital preservation planners thus face a dilemma. Should their efforts focus only on materials readily brought to light by a review of digital objects stored within an institution’s network, at the risk of overlooking materials of
great value? Or, should they plunge into the daunting prospect of attempting to work with individual faculty members in order to include materials of enduring value stored within their collections of devices and accounts in a digital preservation program?

Conclusion

As a part of a larger project exploring digital preservation issues at medium-sized and smaller universities, and potential ways of addressing them, this study sought information describing what types of work-related digital objects professors working at five institutions produced; how they stored and managed such as they considered to be important to their professional activities, and why; what data losses they experienced; and how successful they believed their data-management activities would be in rendering their materials available for future use. In gathering responses from fifty-six participants, the study produced data that are important for a number of reasons. Librarians, archivists and other professionals wrestling with the challenges posed by the ongoing preservation and use of digital objects have largely framed the problem by emphasizing that several factors, including software and hardware incompatibility and imperfect or otherwise compromised storage media, can lead to digital content loss. This study’s findings, by contrast, show that over half of its participants have indeed experienced the loss of at least one digital object that they considered important.

Although these results may initially be discouraging in that they document the ongoing loss of digital materials in a university setting, they are hopeful in that they suggest that these losses are very likely not due to mysterious, unknown causes. Rather, they can be correlated with a set of readily comprehensible and problematic dynamics and practices. The study showed that participating faculty members produced digital objects in a large number of formats and file types. Most were largely ignorant of basic
principles of personal data management, and many sought to manage and preserve them by relying on a wide array of free-standing devices and service-provider accounts, rather than their university network. These risk-enhancing practices can initially be addressed directly and mitigated by faculty members' higher rates of network storage usage. This study showed that a number of participating faculty members were unaware of the availability of network storage or found it inconvenient to use, suggesting how colleges and universities may readily increase faculty members' network usage.

In the long run, the broader implementation of programmatic measures developed to curate materials in keeping with granting agencies’ data management mandates can preserve other faculty-created digital objects even more effectively. Empirical data showing that when university professors were left to their own devices in the management of digital resources over half of them lost the future use of materials that they considered important to their work should alarm faculty members, as well as the deans, provosts, and university presidents ultimately responsible for safeguarding an institution’s intellectual capital. In this frame of mind they may become more likely to assign a high priority to the extension of programmatic digital preservation measures to include selected elements of this type of data as well. Unfortunately, the study’s findings also suggest how faculty members’ personal data management practices can undermine attempts to recognize and include their materials in an institutional digital preservation program. The work of identifying, gathering up and evaluating digital resources scattered across numerous personal storage devices and accounts will require considerable, persistent effort by digital preservation professionals and faculty members alike. When asking faculty members to weigh the costs and benefits of undertaking or promoting such an effort, digital preservation advocates can perhaps benefit from asking them to reflect on their underlying, perhaps unrecognized, assumptions about their
digital objects' viability in the future. A discussion of study respondents' rates of digital data loss and their often-lofty expectations about the future availability of digital materials can perhaps induce different members of a university community to realize that they have been assuming that their materials will exist indefinitely in a readily usable form, while doing little or nothing to insure that they survive into even the near future.
Acknowledgements

The authors wish to thank Lynne Thomas, Stacey Erdman, Katharine White and Sarah Fraser of Northern Illinois University Libraries; Aaisha Haykal of Chicago State University Libraries, Dr. Patrice-Andre Prud'homme of Illinois State University Libraries, Meg Miner of Illinois Wesleyan University Libraries, and Dr. Jeff Hancks of Western Illinois University Libraries for their collection of data for this paper.

References


[Congressional Bill] Fair Access to Science and Technology Research Act of 2013, H.R.


[Journal Article] Liebscher, P., Abels, E., & Denman, D. Factors That Influence the Use of Electronic Networks by Science and Engineering Faculty at Small Institutions II:
Preliminary Use Indicators. *Journal of the American Society for Information Science* 43 (6) DOI: 10.1002/(SICI)1097-4571(199706)48:6<496::AID-ASI3>3.0.CO;2-P


Results from POWRR team interviews with faculty members at the 5 project partner institutions – Fall 2012

Interview Questions:

- Please tell me about the work/research you do and its significance.
  - What kind of data does your work generate?
  - Where do you put this data to keep it safe?

- If you lost all of your electronic files, what are the 3 files you would attempt to recover first and what would your approach be to recovering those files?
  - File category (Research Data, Scholarship, Teaching Materials, Communications, etc.)
  - File location (where it is physically stored...NIU server, external hard drive, Google Docs, another institution’s server, thumb drive, etc.)
  - File format(s) & size
  - File creation date
  - File stewardship
  - Recovery approach

- Have you ever experienced a significant data loss? If yes,
  - Was there University involvement in the attempt to restore your data?
  - What were the repercussions of the data loss?

- Have you experienced an "unrealized" data loss (For instance, where you have data saved in an older software version or on an older piece of equipment like a floppy disc that you probably couldn't access if you tried?

- Are you aware of any formal or informal policies from NIU, your college, or your department that govern how you should be protecting your data from loss?
  - If yes, please elaborate

- Are you bound by a granting agency data management plan?
  - If yes, what is your plan?

- If you were to suffer a data loss, how confident are you that your data could be restored? (Scale of 1 to 5 – 5 is “Very Confident” and 1 is “Not At All Confident”)

- How confident are you that your electronic files will be understandable and usable by your colleagues 25 years from now? (Scale of 1 to 5 – 5 is “Very Confident” and 1 is “Not At All Confident”)

- Do you care if your electronic files will be understandable and usable by your colleagues 25 years from now?

- What have we not asked you about your data that you believe we should know?

- Would you be willing to allow our team to have a copy of one or more of your electronic files to be used in the testing of various Digital Preservation Systems? (please see Informed Consent)
Where do POWRR Faculty keep their data?
What category of files do POWRR Faculty want to restore?

- Research Data: 30
- Teaching Materials: 23
- Administrative: 10
- Communications: 5
- Scholarship: 12
- Organizational: 4
- Personal: 13
- Other: 4
What format of files are POWRR Faculty storing?
Data Loss Experienced By POWRR Faculty

- Yes: 31
- No: 25
Unrealized Data Loss Experienced By POWRR Faculty
Confidence Level of POWRR Faculty that if lost, their data could be restored.
Developing a Cooperative Network for Distributed Digital Preservation

an

Analysis and Model Agreement

preparing

Dwayne K. Butlter, J.D.

Evelyn J. Schneider Endowed Chair for Scholarly Communication
Professor, University Libraries
University of Louisville, Louisville, KY USA

© 2014 - Dwayne K. Butlter

---

1 This model agreement presents a general framework for developing a cooperative membership organization. The purpose of model legal agreements is to identify broad legal issues and suggest ways for managing individual legal concerns, including here the organizational and legal structures supporting digital preservation cooperatives. A host of varying laws and other considerations would govern the planning and development of an actual cooperative agreement and enforcement of any contract used in specific known legal jurisdictions. Model agreements are not intended to substitute for diligent inquiry and individual legal advice. As a result, the analysis and model agreement does not constitute legal advice. Readers planning the development of their own cooperatives should consult their own counsel in order to understand and manage effectively their own legal needs.
HOW TO USE THIS ANALYSIS AND MODEL AGREEMENT

This analysis and model agreement centers on its two distinct approaches to addressing and supporting the underlying goal — developing a legal agreement governing a membership cooperative. Some of the text can serve as model language for developing the cooperative agreement, provided that after close reading the language accurately reflects the individual needs and goals of the planned cooperative. This model language is broad, factually nonspecific, and generally applicable to not only common legal concerns arising in many jurisdictions but also to the issues and situations described within the model language. The model language is labelled “Model Language” and typically follows a section number directly to aid in identification.

In the alternative, some of the text is pure narrative and strives to identify and describe some common themes, needs, and possible solutions to developing a digital preservation cooperative. The narrative is purposefully broad and intended to give cooperative planners a foundation for recognizing how those themes, needs, and possible solutions connect to and interact with cooperative planning and ultimately developing the membership agreement. The narrative sections may appear obvious through context, labelled as “Principles and Goals,” appear as bracketed “NOTE(s),” or identified as “Example(s)” in order to distinguish them from model language sections. The role of the narrative is ultimately to give cooperative planners choices and alternative approaches to resolving the issues and legal concerns unique to their cooperatives. Knowing alternative approaches provides the support for drafting an agreement and language that not only embraces individual and unique concerns but also encourages managing them effectively. Not surprisingly, these cooperative-specific sections within legal agreements are inevitably fact specific and require a clear understanding of the real-world cooperative structure that they will govern. These sections also would typically reveal jurisdictional variations within law and will require close analysis in order to understand their local and unique implications for the membership agreement.

The Makings of An Agreement – Developing the Cooperative

The discovery and balancing of costs, rights, and responsibilities for developing a digital preservation cooperative and supporting network requires an analysis of some fundamental structural and organizational questions. Parsing these questions can help the cooperative planners identify key issues and frame appropriate administrative, governance, and legal responses for

---

2 Legal professionals and courts sometimes use the term “agreement” to indicate that “the agreement” is not in fact a legally enforceable “contract.” That usage of “agreement” essentially means that the relationships and condition embodied in “the agreement” fail to satisfy the basic legal requirements for forming and enforcing the “agreement” as a “contract.” In that usage, “agreement” means an informal and unenforceable instrument governing an arrangement or relationship. On the other hand, “agreement” as used in this analysis is synonymous with “enforceable contract.” This usage more accurately reflects the general lay usage of “agreement” to mean “enforceable contract.”
managing them. Making these choices and resolving those issues creates the foundation for drafting a membership agreement that accurately reflects the choices and final preferences and provides the framework for managing them. That early decision making is also the cornerstone for crafting governance, administration, and the form and makeup of the digital network.

Many of these framing questions implicate an equally important and more central organizational question: should cooperative planners address and manage the issues and preferences at hand as contractual terms, governance and/or administrative rules, or in technical or other documentation necessary for governing daily operations? This list of structural and organizational questions can help frame planning and approach, but it is by no means exhaustive and not surprisingly does not include all issues that could arise in developing a digital preservation cooperative or similar collaborative organization.

**Structural Issues:**

1) Who are (or can be) members of the cooperative? What would the cooperative look like from the structural and governance perspective? Who initiates and conducts the planning? Is there an institutional host willing to undertake business activities requiring a legal existence or to incorporate in order to support them? What will the membership agreement include and what positions would it take?

2) What would the governance look like? Given the nature of collaborative efforts and cooperatives, balancing the interests of administrative necessity, central control, technical and structural interoperability, and shared understandings and perspective on goals and aspirations can require careful political maintenance and management. How would a central administration and preservation nodes allocate those and other responsibilities while reporting to possibly different constituencies and likely sharing different administrative rules, requirements, and conditions? How could the cooperative consolidate those responsibilities and duties? Will the cooperative exist under a larger umbrella organization or institution?

3) Who makes membership decisions? Membership committees applying a governance charter or set of operational bylaws? A cooperative membership agreement setting forth specific
criteria? A host institution such as nonprofit, university, or something else? An executive
director or other administrator reporting to a board of directors or other oversight group?

4) Who defines member and creates membership categories and/or tiers? Who sets fees and
any service charges? A steering committee, or other administrative or participatory group? How
are the fees set? What are the financial goals? Sustainability or more? Are those fees sustainable
or subsidized in some way by a larger organization or grant funding? Who applies for and
manages financial and/or grant requirements?

5) Who researches and reviews technology and makes platform, storage, and support
decisions on behalf of the cooperative? How is supporting technology acquired? Institution?
Cooperative? Other organization?

6) Would cooperative members have equal voting power? Proportional voting? Categorical
voting? Would each institution, a bloc, or central administration have a vote or veto over certain
opportunities or assignments occur through appointment, election, or membership? What form
would best serve the cooperative? Individual members? What benefits would accrue to members
that choose to host a node and undertake research and development relative to members who
solely require preservation services?

7) Is there a strategy for coordinating administrative services if collaborators or the
cooperative make up changes membership or hosts? What sustainability models can support the
cooperative if central administration changes through time or dissolves completely? Can one
institution take over the work of the cooperative in the event of shifting priorities? Is there a
succession plan to manage potential changes in the makeup of key principals or members of the
cooperative?

9) Would the cooperative or an institution serve as the legal entity/representative for the
cooperative in order to manage potential grant funding opportunities, problems arising among
hardware vendors or other collaborators, insurability concerns, compliance with external
influences, rules, and regulations?
10) What is the relationship among members relative to the central preservation node and the administrative entity if different? The same? How can the cooperative manage member relationship and possible tensions between them? How can the cooperative allocate time and resources relative to members or needs of cooperative?

**Organizational Issues:**

1) Who will serve as lead contact(s) for the cooperative regarding administrative, financial, and technical questions? What the reporting lines for cooperative personnel or members holding assigned or specified duties within the preservation network and/or the cooperative? How are in-kind donations of services or programing applied to the preservation network assessed or valued institutionally and within the cooperative? Who owns the network and/or hardware components in it? Will software development or other shared work produced intellectual property of interest to the respective institutional members? How would cooperative resolves such issues? Can entities agree to common licenses to share developments? How will the cooperative and members manage copyright and other legal concerns arising within the content and resulting from the reproduction and distribution of copyrighted works across multiple nodes for preservation purposes? Is the archive “dark”? Why is it “dark”? What are the implications of opening all or part?

2) Would the cooperative provide network storage or secure third party storage while the cooperative manages network operations? What would a service storage agreement look like and who would approve it? Who assumes liability, if any, for errors in storage of content or operation of the storage facility? Does the cooperative act as an agent for each member or is each member directly bound to the storage agreement and/or other agreements? Does the membership agreement requires agency appointments for members in order for the cooperative to pursue cooperative activities benefitting members and the cooperative?

3) Can an assigned committee oversee day-to-day network operations and problems and adequately address ongoing administrative, billing, organizational, and technical concerns or would that authority rest in an Executive Director and central project personnel reporting to an
oversight/advisory board for the cooperative? What types of agreements are necessary in order to manage these relationship and to protect the interests of members, the cooperative, and central administration? Who drafts these agreements? Under whose direction?

4) Would each collaborative member hold liability for its own contributions and work within the network or would a joint liability umbrella exist within the preservation nodes or in the cooperative? How would participating institutions and/or members manage individual and possible joint liability? Is indemnification of the cooperative necessary and under what conditions? Will state entities that join the cooperative provide indemnification for the cooperative? How will the cooperative manage individual requests to change the agreement? Change fees? What tracking mechanism would track these changes and any deviation from the standard agreement? What are the administrative implications of allowing changes? Denying them?

5) How will members withdraw from the network? What happens to their content? What if the member operates a preservation center or node? What if the content is managed on behalf of a “sponsored” member of some type that cannot or does not pay fees? Can the collaborative prohibit further access by the members under certain conditions such as breach of the agreement, network security, or legal challenges to the content? Does the contract become a “trust” in any sense? Does the cooperative hold any obligations to the members beyond the preservation missions? Is there a fiduciary relationship? Are there local, federal, or international laws governing the sharing of content? Does the content contain personally identifiable information? Will the cooperative operate in other countries, the European Union, or elsewhere? How will the cooperative manage confidential information and/or personally identifiable information that may produce and/or acquire through its routine operations? Who owns metadata or similar data and/or potentially valuable information produced under the auspices of the cooperative?

6) Who provides marketing support for the preservation network? Education? Outreach? How can the cooperative make its services widely known and competitively offered? How are those activities funded? Central cooperative or members?
7) What level of technical support and storage space is necessary in order to work within preexisting networks? Will the cooperative build a new network? Are the technical support positions at the intuitional level new positions or apportioned assignments adding to the existing duties of an institutional employee? Who is responsible for collaborators who only contribute content but no technical know-how or support? Who provides expertise or technical services to such collaborators? How will the network define custodial access for network security or to satisfy other copyright or legal requirements? Who would qualify for limited access as a custodian? How would rights management architecture work as implemented across potential multiple systems architecture? How is custodial access managed and tracked?

8) What strategies are in place for assessing various organizational risks and compliance requirements of respective organizations? Supporting institutions? The cooperative? Who serves as coordinator for those administrative concerns? How can the membership agreement best allocate risk among members and between the cooperative and its members? How will the cooperative evaluate appropriate and necessary risk relative to the legal and technical landscapes?

9) Who oversees legal issues relating to the work of the cooperative? Who provides legal counsel to the cooperative and/or members or other collaborators in order to assess legal risks of included content or network security risks? How risk tolerant is the cooperative? The members? Umbrella organizations or institutions? Are there risk management considerations at the institutional level, the cooperative level, and/or within collaborating institutions? Are any of the efforts grant supported? How are grants administered relative to cooperative and members? Do any grant conditions define or influence the legal, administrative, or technical architecture of the cooperative?

10) How can the cooperative affiliate with other institutions in order to achieve specific goals? What would affiliation look like and what kind of legal agreement would best reflect affiliation and the mutual obligations and benefits of the relationship?

These breadth of these structural and organizational questions underscores the need for detailed planning in developing a cooperative, but they are no substitute for exercising diligence in
understanding and managing the range of possible implications and concerns relating to the cooperative. On the other hand, these questions offer key support for framing the next task of dissecting the form and substance of a model membership agreement in order to show how cooperative developers might draft an agreement otherwise specific to meet their needs and possibilities. Equally important, the framing questions can also help convey to their local counsel numerous alternatives for drafting an agreement that distinctly reflects the culture and mission of preservationists, archivists, and librarians and reinforces the need for trust and flexibility inherent in developing a digital preservation network.

**Model Agreement – Possibilities, Choices, and Structures**

Contrary to common speculation and sometimes experience, the desired goal in creating any legal agreement is precision and clarity. Precise language and well-conceived structures will reinforce that clarity and provide a roadmap for reading and understanding the agreement, benefiting all parties. Cooperative membership agreements are no different. The central concern is whether the cooperative and members fully understand and can agree to the terms and conditions within the membership agreement. Fully understanding and identifying commitments in advance lessens the likelihood of misunderstandings afterwards. Digital preservation is a complex undertaking requiring planning, people, information and digital technologies, and ultimately trust in relations and the sharing of a common goal. A clear agreement can help reinforce that approach and set the tone for positive member relations.

The following “model agreement” will consist mainly of further narrative descriptions and guides for identifying and interpreting the underlying issues relevant to individual sections and legal concerns in order to begin drafting language and terms to describe and respond to them. However, some of the sections will provide precise language that could prove useful in some agreements with little further modification other than capturing and then reflecting individual local needs or concerns, characteristics exclusive to a specific cooperative, or local laws or administrative considerations sharing those unique qualities.
Model Language

[CAPTION] [Insert Name of Cooperative] Membership Agreement

1. The purpose of this agreement

This membership agreement identifies and defines the relationships among members of the digital preservation cooperative. This introductory overview offers a general guide to understanding the expectations of the members as well as recognizing the requirements and responsibilities of each member and the broader cooperative. The overview constitutes an essential part of the entire agreement and is read as a part of the entire agreement in order to fully understand all requirements, responsibilities, terms, and conditions of the entire agreement.

The purpose of the cooperative agreement is to translate the general aspirations and structural expectations that initially motivated the cooperative planning from tentative concepts to ordered legal relationships. Those relationships and their legal foundation support not only the work of the cooperative but more importantly embody and protect the legal interests of the cooperative and all its members.

To that end, this membership agreement establishes an enforceable contract that describes the cooperative, defines its membership, reflects its governance structure, and orders the relationships within the cooperative and among members. The detailed provisions and specific language require close reading and ultimately bind the cooperative and each member legally to understanding and fully undertaking and discharging their responsibilities and obligations under the agreement accordingly.

The central foundation of agreement among all members is that digital preservation technologies, tools, and strategies represent a new and complex challenge in undertaking and managing preservation in the digital realm and as such lack sufficient study and reliability to warrant the long-term application of them or the enduring preservation of any particular objects or items at any particular time. The discipline of digital preservation is still in infancy and maturing slowly with experimentation and diversions in the technological, strategic, political, and sustainability
realms. Consequently the cooperative and/or individual members specifically and fully disclaim any liabilities and/or real or potential warranties assuring the successful and/or persistent operation of the cooperative and digital preservation activities. Members agree to waive any liability claims in in legal context that arise against the cooperative in the course of preserving content and/or managing, administering, and/or planning preservation activities and strategies.

2. Meaning of the Agreement and Interpretation

Trust among members and the cooperative provide the compelling rationale and support for accomplishing the key and essential but intricate tasks of digital preservation across multiple institutions with multiple rules and multiple perspectives. Much like digital preservation and the cooperative itself, the membership agreement reflects a living and evolving set of principles, strategies, and relationships that are precisely defined in order to achieve clarity and an important shared goal — long-term digital preservation. Consequently, like the underlying discipline, the membership agreement requires equal vigilance to stay current and accurately reflect not only the underlying trust of members but also the living and evolving nature of digital preservation as a new discipline and practice. Revisiting and revising the agreement as necessary requires redefining terms and relationships as necessary and subject to governance structures and conditions in the agreement and cooperative structure.

To that end, members as a core requirement and condition of membership agree that flexibility and modification are equally important to the strict application of this membership agreement in ordering the respective relationships of the cooperative and in manifesting the intent of the parties that cooperate within the cooperative.

3. Description of the Cooperative and Technological Considerations

**Principles And Goals:**
This section provides an essential opportunity to identify your individual cooperative by name, explain its goals, structures, and relationships with other entities and technology, and precisely identify the planned supporting technologies used to preserve content and support related
activities. In essence, this section should frame for readers not only how and why they can or should join the cooperative (i.e. benefits) but most importantly define any thresholds for joining and precisely explain and bind them to satisfying structural, technological, financial, and legal conditions for joining. Given that each cooperative has nearly infinite choice in creating and pursuing digital preservation, the cooperative or its counsel can only fully develop these terms and conditions after cooperative planning and structuring is complete or nearly complete. By definition, this important factual limitation precludes drafting any “model language” in advance of knowing the facts and creating the cooperative.

4. Definitions for Understanding and Interpreting this Agreement

**Principles And Goals:**
Creating a list of “definitions” is essential in the cooperative membership agreement. This section will define key words and concepts and lay the foundation for framing and clarifying their meaning within the agreement and between the cooperative and its members. Moreover, the definitions agreed to by the cooperative and the members will provide legal readers, including courts and institutional counsel, a set of terms that accurately reflect the understanding and agreement of the members within the context of the agreement.

Agreements sometimes include numerous and sometimes few definitions. As a general principle, the more complex the agreement and/or the less “ordinary” the words or concepts, the more likely that the agreement author(s) should define them within the agreement. However, there is no fixed rule governing what the agreement should include in its definitions section. Some guiding principles would include when the word or concept is ambiguous on its face, needs clarification in context, and/or reflects a cultural, social, technological, or atypical usage, then defining that work or concept is probably necessary and usually helpful.

A useful, albeit not exhaustive, set of definitions might include the following terms and concepts.
Example: 3

"Agreement" means this [INSERT COOPERATIVE NAME] Membership Agreement and any other documents identified herein and specifically incorporated by reference.  

[NOTE: “Incorporate by reference” is a technique that allows lawyers to reference other documents from within an agreement in order to make them part of that agreement. These referenced documents usually contain factual information or related and supporting terms and conditions. One strategic and practical consideration is where to place certain requirements and conditions – in the agreement itself or in documents incorporated by reference into it. Resolving that question occurs in part by determining to what extent the parties want to revise the agreement and/or supporting documents through time. In general, technology standards and administrative rules tend to require revision more often than the basic terms and conditions of the broader agreement. As a result, incorporating them by reference into the broader agreement may prove more efficient than revising the overall agreement each time minor changes in technology or administrative support occur. ]

“Charter” [OR GOVERNANCE DOCUMENT] means [INSERT NAME OF DOCUMENT] as duly amended from time to time. The Charter established the core membership rules and allocation of membership voting structure, rights, and relationships, including establishing membership categories and criteria.

“Collection” means the aggregated content to be preserved under the banner of one metadata record. [NOTE: This definition would become agreement specific and embody technological approaches and capabilities of the particular cooperative. Collection might, for instance, also include the entire corpus of work in some agreements. That outcome would then require defining lesser units in some way so that the members have a common vocabulary for descriptive and

---

3 Many of the examples within this analysis come from the MetaArchive Membership Agreement. The MetaArchive is an early and ongoing distributed digital preservation cooperative that initially formed in 2004 at Emory University with funding support from the National Digital Information Infrastructure and Preservation Program (NDIIPP). The MetaArchive Cooperative subsequently became self-sustaining as a part of the Eduopia Institute, which was incorporated in part to provide a legal identity to support the work of the MetaArchive. This author was the central architect and drafter of the MetaArchive Membership Agreement, http://www.metaarchive.org/public/resources/charter_member/ma_2014memberagreement.pdf. That agreement has served as a widely-regarded framework for national and international efforts to produce the legal foundation for digital preservation cooperatives. Additional information about MetaArchive, including its governance structure, is available at http://www.metaarchive.org/ (last visited November 22, 2014).
legal purposes. The goal is always carefully and accurately capturing the meaning within the underlying concept.]

"Content" means any of the digital materials submitted by Members and satisfying the required collection, legal, and technical standards established by the Cooperative, including any copies of the materials as submitted and those subsequently migrated to meet new format or technical standards.

"Cooperative" means the Cooperative and its Members in good standing.

"Dark Archive" means that access to content is strictly limited within the cooperative to organizational custodians who are clearly identified and meet certain administrative requirements of the agreement or documents incorporated by reference.

"License" means all those rights necessary for the Cooperative to work with all Members in preserving content through a Preservation Network that will create multiple copies, distributions, and possible derivatives of a work, including public performance and display rights. [NOTE: License needs and questions will vary depending upon jurisdiction, governing law, goals of the cooperative, and exposure and legal risk across the network.]

"Material Breach" means any event that violates the terms of the Agreement or governing documents including, but not limited to these possible conditions of breach:

Nonpayment of fees or failure to provide one-year notice of cancellation;

Failure to comply with technical or communication standards or governance and collaboration obligations identified in the membership agreement or Charter;

Copying and distributing content via or from the Network that infringes copyright or other activities that violate other local, state, federal, or international law; and/or

Members also may include former Members and potential candidates for membership in some contexts of the Membership Agreement in order to bring meaning to the intent of this Agreement and to satisfy any provisions surviving the termination of this Agreement through fault or otherwise.

“Network” indicates the technical network and infrastructure of the Cooperative. It is comprised of preservation sites administered in geographically dispersed locations by Member organizations (at the sustaining and preservation membership levels). It stores content submitted by Members.

“Technical Specifications” means the current documented recommendations and requirements for participating in the Cooperative and accessing the Network, including staffing and hardware.

“Preservation Site” refers to a site that houses [NOTE: This language should describe the technology supporting the archive, network, and nodes depending upon applicability.]

“We” and “Our” and “Us” means the Cooperative and [POSSIBLY THE BROADER CONTROLLING INSTITUTION] and its legal representative but would not extend to individual officers and directors of the Cooperative or supporting institutions for liability purposes.

“You” and “Your” and “My” means the entity seeking membership in the MetaArchive Cooperative and those entities who achieve membership, including more specifically for agreement and signature purposes, a representative of that entity who holds signature and agency authority to create agreements binding that entity.

In some sense, the definition list cannot include too many definitions but it can contain too few. A helpful balancing test to determine what to include is whether the value of the necessary detail outweighs the value of satisfactorily achieving overall brevity and clarity and while simultaneously best protecting the legal interests of the cooperative and members. On the other
hand, if the cooperative planners hold any doubt as to the meaning of individual terms, then defining them within “definitions” is a prudent choice.

Much of the misunderstanding that arises in contract law comes from differing perspectives on the meaning of specific terms and phrases. To some degree, this outcome is inevitable because no language is amenable to absolute definition. However, making a deliberate effort to define important terms or ambiguous concepts will not only help provide greater legal certainty but also will offer reassurance to all parties that they understand the relationship and can move forward with a high degree of predictability and shared trust.

5. Overview of Membership and Fees

Principles And Goals:
Carefully describing the criteria for membership, eligibility, fees and related charges, rights, and concordant responsibilities is crucial to drafting the membership agreement. Those criteria will consist of what are unsurprisingly local and cooperative-specific needs, goals, and considerations. Crafting the member relationships between members and the cooperative delineates not only how the cooperative will function day-to-day but also provides the foundation for assessing how governance could work, including outlining necessary committees, the duties of “central administration” such as procuring 3rd party services, billing, coordinating education and outreach, managing membership agreements and requests, troubleshooting and coordinating technology and technology strategies, and setting prices and any related charges. Isolating those relationships and structures also encourages close scrutiny of the technological goals of the cooperative, the possible architecture for achieving them, and how the members, committees, and cooperative could divide and share responsibilities and functions across the cooperative. Equally important, this close scrutiny and planning will provide the framework for identifying and defining what types or categories of membership would best serve the shared goals of members and the cooperative.

Membership categories or tiers can serve different needs and goals. Underlying the creation of them and therefore reflected in the membership agreement and its relationships, rights, and responsibilities are core issues and decisions that are necessary to make in order to develop the
cooperative and then craft the membership agreement. These core decisions include structural, organizational, legal considerations, and many choices.

Distributed digital preservation, for instance, requires a robust network and multiple network nodes in order to assure long-term preservation. Those nodes are otherwise insulated from geographical or other similarly disruptive calamities. Hosting a node requires investment in a technology infrastructure and likely would entail a more rigorous time and resource commitment to the cooperative. Not surprisingly, tinkering under the technological hood holds greater benefit for the cooperative, but holding responsibility for network operations also requires greater time and resource commitments from the members that operate nodes. Conversely, some members may prefer no tinkering and simply want to cost effectively preserve their collections with minimal time and resource investment. Those aspirations and goals also have a cost that factors into creating membership types and allocating costs and risks across the cooperating and among its members. The MetaArchive Membership Agreement offers some context for how other cooperatives have addressed the central architecture of membership.

Example:

The MetaArchive Cooperative offers membership opportunities in three distinct tiers relative to the goals and preservation needs of Members. [NOTE: Some cooperatives may want to create fewer or more membership categories depending on the unique needs of that cooperative. As a general rule, carefully identifying membership goals and possible categories earlier in the planning process is helpful in framing the rights and responsibilities of each category relative to other categories. Assessing the incentives for joining one versus another category is crucial to understanding other related structural and organizational questions, including how work is apportioned and how to shape the network in ways taking a maximum advantage of each members unique strengths and capabilities.]

Sustaining Membership
Sustaining Members are integral to the research, development, and deployment work of the MetaArchive Cooperative and agree to contribute staff and resources to the effort and to participate as addressed more fully under Member responsibilities (below and in the
MetaArchive Cooperative Charter document). The fee for joining as a Sustaining Member is $5,500/year for a three-year period with annual billing opportunities. Sustaining Members that pay in full for their three-year period during their first six months of membership will earn a 10% deduction and will pay $14,850 for their three-year membership fees. Space is purchased from the Cooperative on an annual basis, and the 2012-2013 rate is $1/GB/year, which covers hard storage costs and administrative overhead involved in ingesting and preserving content.

Preservation Membership
Preservation Members engage in ongoing preservation activities and agree to participate in the network including maintaining a MetaArchive-LOCKSS cache and other responsibilities set more fully under Member responsibilities (below and in the MetaArchive Cooperative Charter document). The fee for joining as a Preservation Member is $3,000/year for a three-year period with annual billing opportunities. Space is purchased from the Cooperative on an annual basis, and the 2012-2013 rate is $1/GB/year, which covers hard storage costs and administrative overhead involved in ingesting and preserving content.

Collaborative Membership
A Collaborative Member is a group of institutions that run a shared, centralized repository and contribute this shared content to the MetaArchive Preservation Network (please see eligibility requirements as outlined in the MetaArchive Cooperative Charter). Collaborative Members engage in ongoing preservation activities and agree to participate in the MetaArchive network including maintaining a MetaArchive-LOCKSS cache and other responsibilities set more fully under Member responsibilities (below and in the MetaArchive Cooperative Charter document). The fee for joining as a Collaborative Member is $2,500/year for the collaborative, plus a small fee per year for each participating institution.4 Membership is for a three-year period with annual billing opportunities. Space is purchased from the Cooperative on an annual basis, and the 2012-2013 rate is $1/GB/year, which covers hard storage costs and the administrative overhead involved in ingesting and preserving content.

Collaborative Members occupy a unique role in the MetaArchive Cooperative and therefore further agree that Collaborative Members will assume sole and singular legal and other

---

4 Rates are negotiated with the Program Director and are based on the size of the collaborative, and will not exceed $100/institution/year.
responsibility for the included content and related actions of users whether Collaborative Member has or has not secured individual agreements or contractual relationship with contributors to the Collaborative. Collaborative Members also acknowledge and agree that the Collaborative Member will set in place appropriate administrative and policy arrangements in order to assess the copyright, intellectual property, and/or other legal implications of contributor’s activities relating to the Collaborative arrangement.

Fiscal payments will be accepted between January 1 and December 31 of each year in order to allow each Member to schedule such payments consistent with its own fiscal schedule. At the time an organization joins, it must alert the MetaArchive of the date upon which the Member intends to submit payment to the Cooperative (see the Signatures Page).

Some rights and responsibilities vary across tiers consistent with the expectations and benefits available to each tier and the fees administered for joining that tier. We reserve the right to be flexible with regards to each membership category’s rights and responsibilities.

All Members, regardless of their classification, share some rights and responsibilities in common, including agreeing to membership terms of no less than a three-year duration with a one-year notice to cancel membership. This time frame provides some predictability for the Cooperative as it continues to grow and will help improve services to all Members by lessening administrative burdens and thus potentially lessening membership fees. The Cooperative will address failures to provide appropriate notice according to the MetaArchive Cooperative Charter and rules governing continued access to materials and breach set forth elsewhere in this MetaArchive Membership Agreement.

The MetaArchive Cooperative offers other services to all membership levels on a consulting basis, including consulting and training. Such services may be engaged by negotiating a separate Service Agreement with the MetaArchive Cooperative.

Joint MetaArchive/NDLTD Membership Discounts: Organizations that simultaneously join the MetaArchive Cooperative and the Networked Digital Library of Theses and Dissertations (NDLTD) are eligible for discounted membership fees. In the case of the MetaArchive Cooperative, membership fees are discounted by 5%. Organizations interested in joint
membership discounts should contact both the MetaArchive Cooperative and the NDLTD to confirm their eligibility before completing this form.

5.1. Your Rights as Members

Principles And Goals:
To some extent, the central role of contract law and enforceable agreements is to ensure that transactions across the range of human enterprise appropriately allocate risk and define the rights and responsibilities of each signatory to the agreement. It is through this lens that legal practitioners and courts view the agreement and its core purpose. Rights and responsibilities set the foundation for understanding what the parties want (or must have wanted) and what they were willing to give in return to resolve a shared concern, common interest, or mutual goal.

Carefully identifying and then defining rights and responsibilities is key to producing a successful agreement and ultimately a successful cooperative. The rights and responsibilities section sets the tone for member relationships and signals how the cooperative will manage them. The rights and responsibilities describe the benefits of joining the cooperative and obligations for making it work effectively. Building a comprehensive platform for technological support and governing the cooperative is an essential step taken prior to drafting the membership agreement. That step allows planners to fully realize the scope of the rights and responsibilities as between the members and cooperative and among all members. Not surprisingly, given the collaborative nature of cooperatives, unlike some membership arrangements, those agreements governing cooperatives also sometimes include rights and responsibilities that flow through the cooperative and influence all members individually and collectively.

These flow-through possibilities also require careful scrutiny. Contributing copyrighted content, for instance, could raise not only contributory liability issues for the cooperative but also direct liability for other members because the copyrighted content is replicated through the network and across multiple preservation nodes. As a result, each preservation node also potentially holds direct liability for infringing activity simply by reproducing and distributing the content as part of its preservation obligation. In this context, indemnification (i.e. paying the fees for the
cooperative to defend itself from liabilities produced by the acts of members), while sometimes problematic in some agreements, can prove immensely useful in a cooperative agreement. An indemnification requirement can produce incentives for all members to carefully consider their own practices regarding copyright and other potential flow-through liabilities. A member might, for instance, choose judiciously to prevent causing liability within the cooperative not only to avoid indemnification obligations but also to prevent flow-through liability to that or other members. Each member’s self-interest can benefit the cooperative and each other member by encouraging legal compliance and engendering trust in the work of other members.

Right and responsibilities in particular will prove unique to each cooperative given the nature of developing the cooperative and creating the scope of combined effort in order to accomplish preservation. To that end, the structure of MetaArchive agreement is instructive but not determinative regarding the rights and responsibilities arising in other cooperatives.

Example:

Your rights as a Members:

Each Member working with the [MetaArchive] Cooperative receives:

Distributed archiving of contributed digital collections across multiple, geographically distributed preservation sites;

Retrieval of the Member’s content in case of a catastrophic loss;

Assistance with the installation and maintenance of MetaArchive-LOCKSS software, documentation of processes and technical standards, and technical support;

Reports about the Member’s submitted content and about the overall Preservation Network;

Service opportunities within MetaArchive’s Committees (Content, Preservation, Technical);

Attendance and participation at Cooperative conferences and workshops at discounted rates;
Opportunities to collaborate with and/or learn from experienced digital preservation administrators, librarians, technologists, and others who work with the Cooperative;

In the case of catastrophic circumstances, the ability to request technical and financial assistance with the restoration of a preservation site’s caches, software, and collections by the MetaArchive Cooperative. These requests will be reviewed and, at the discretion of the Steering Committee, either approved or denied;

Access to the technical knowledge and expertise of Cooperative Members and technical support to establish and maintain preservation sites in compliance with the MetaArchive Technical Specifications; and
Additional preservation services that can be purchased at a contract rate (e.g., consulting and training around preservation issues).

Sustaining Members also receive:

The ability to influence the development of the MetaArchive Cooperative and the development of distributed digital preservation practices more generally; and

Voting representation on the Steering Committee (see 3.2.2 of the MetaArchive Cooperative Charter for more detail).

Collaborative Members also receive:

The ability to offset the cost of preservation by capitalizing on their existing, collaboratively hosted repository infrastructures.

**Your Responsibilities as Members**
All Members working with the Cooperative agree to:

Maintain membership in good standing by fully complying with this MetaArchive Membership Agreement and the definition of herein and by acknowledging and agreeing to processes,
procedures, and standards of governance found in the MetaArchive Cooperative Charter and with technical requirements identified in the Charter or other technical documents identified in Cooperative publications or notices;

Create and maintain a MetaArchive-LOCKSS cache that complies with all requirements in the current and subsequent versions of the Private LOCKSS Network software and make this cache available for testing new software and other Cooperative developments as needed;

Design and implement system features ensuring compliance with Cooperative security requirements and content validation, including but not limited to integrity checking as well as metadata analysis and tracking;
Install and maintain any other software that may be required for participation in the Cooperative;

Participate actively in the MetaArchive Preservation Network by ingesting and monitoring content at the Cooperative’s request;

Reproduce and distribute content submitted by other Members only to the extent necessary to conform to the requirements and conditions set forth within this MetaArchive Membership Agreement;

Support at your own expense any and all costs incurred by participating in the Cooperative, including but not limited to paying membership fees, travel to required meetings, and other costs of participating in the Cooperative;

Implement and deploy appropriate standards, processes, procedures, and safeguards for resolving copyright and other legal requirements that might arise in contributing and/or hosting content in order to comply with local, state, federal, and international law, including the use of exceptions set forth within U.S. copyright law such as, but not limited to, good faith, reasonable applications of fair use in Section 107, relevant library reproduction and distribution in Section 108, and elsewhere in the statutes or amendments and/or permissions through “deeds of gift” or other instruments, documents, permissions, or clearance arrangements. Other potential governing laws for Members may require the application of principles such as, but not limited to, country of
national origin governing the inclusion of materials and relevant exceptions and conditions

governing the hosting of materials and resulting specifically from the geographical location of
the preservation node;

Represent and warrant that to the best of their knowledge the Member is not contributing content
to the Preservation Network that would infringe the rights of others and that the Member holds
sufficient rights to License the Cooperative and Members sufficient rights to use the content
consistent with the terms of this MetaArchive Membership Agreement;

Hold the Cooperative and other Members harmless in the event of infringement, claims of
infringement, loss of data, interoperability, and any other technical standards and governance
claims by waiving any rights of recovery for any costs or damages associated with your
relationships and Agreement with the Cooperative;

Indemnify the Cooperative to the extent permitted by law for any losses and costs incurred by the
Cooperative and Members such as but not limited to legal fees, costs, and damage awards arising
from infringement or other claims directly related to your activities in working with the
Cooperative and Members;

Cure any material breaches of the contract within a 90-day period unless the Cooperative agrees
in writing to an extension of the cure period; and

Provide technical and administrative contact information as necessary to enable communication
with other Cooperative participants as needed or upon request by the Cooperative.

Sustaining Members further agree to:

Develop and test hardware and software, network connectivity, and transmission standards;

Engage in the technical development and testing of the computer systems that enable the
MetaArchive Preservation Network; and
6. What if your Membership Ends?

**Principles And Goals:**
Framing a strategy for managing membership withdrawal requires careful analysis. One question considers how to balance the long-term interest of preserving at-risk content for all as a broad social goal against the possible free-rider member that preserves its content as a member and then withdraws to avoid further costs for the preservation. The free rider benefits from the work of the cooperative without providing proportionate funding or resources to support that work. To put it another way, how can the cooperative discourage the act of an entity joining to preserve content and once accomplished leaving fully knowing that the content is in fact preserved for future generations? Discouraging the possibility of free-riding can occur through disclaiming any further effort to preserve content once the members withdraws. The downside of that approach is that it equally lessens the likelihood of the enduring availability of that content or of any content from any member that subsequently withdraws. The legal response serves the goals of the individual cooperative but lessens the success of the broader social goal of widespread preservation.

An equally vexing problem arises in crafting the membership agreement to address possible free riding and to manage other broader social policy expectations. For instance, the general perspective of archivists and preservationists is that long-term preservation is essential for preserving history, culture, and other reflections of human behavior for future researchers and scholars. However, given the technological uncertainty and newness of digital preservation and the ongoing struggles to achieve sustainability within the limits of slim public funding for what is essentially a public benefit, the agreement must state unequivocally from the outset that there are no guarantees from the legal perspective. That positioning while necessary can require careful conversation within the relevant communities of preservation in order to separate the legal rationale for taking that position from the common presumption that risk-averseness permeates the legal community and its decision making.
Achieving the goal of effectively managing member withdrawal can take many forms and involve multiple layers of involvement. The following example is brief and pointed to lessen the overall complexity of the agreement and to leave no uncertainty about its meaning. Other possibilities may include conveying the underlying legal rationale or describing circumstances of how possibilities and situations may differ across membership categories or tiers. Much of that framing will come from how the cooperative structures fees, costs, ancillary services, buy-out possibilities for long-term storage, and balancing the interests among members and the cooperative.

Example:

The Cooperative is not responsible for preserving your content or making it available to you beyond your membership period.

7. Our Rights and Responsibilities as a Cooperative

Principles And Goals:
Crafting the rights and responsibilities of the cooperative also requires careful consideration and provides the second pillar of the overall agreement. Like the member rights and responsibilities sections, understanding cooperative right and responsibilities and their possible interpretation and implications is fundamental to understanding the agreements and the relationship between the cooperative and members. This section essentially defines the balance of power in the cooperative and among members.

Some of the lessons-learned in cooperatives *writ large* generally suggest that central power provides the foundation for the cooperative to work. However, appropriate checks-and-balances must temper that central power in order to establish the trust necessary to encourage a common commitment to its outcome. A member of an electrical cooperative, for instance, actually holds little sway over daily operations other than electing a voting representative each year that presumably would reflect the broader interests of the members and temper central power within
the cooperative. However, interests of individual members are typically diverse attenuating the power of the individual and the likelihood that the elected representative can truly represent all members. Members who feel unrepresented can come to distrust the cooperative and its decisions making. Consequently, the attenuation of power discourages building the level of trust necessary for making a resource-challenged digital preservation cooperative work effectively. It essentially produces a utility-customer relationship notwithstanding the trappings of a cooperative and shared governance. Finding the right balance of power is necessary but require frank conversation and careful planning of the cooperative and the membership agreement.

The governance document is of course key in defining relative powers, but it is the agreement that vests them legally. To some extent, the governance document is a precursor to the membership agreement. Drafting the membership agreement without it is problematic at best and most likely impossible at worst. To that end, creating the governance document needs to occur before drafting the membership agreement in order to describe the balance of power between the cooperative and members and to successfully reach a consensus on the most meaningful structure for serving local and unique needs and expectations.

The rights and responsibilities of the cooperative also should embody the expectations of members and reflect the needs of day-to-day administration and accomplishment in preservation. That outcome seems easy but comes only with careful planning and discussion. In some sense, the rights and responsibilities section for the cooperative reflects the core concerns of not only assigning responsibility for tasks and accomplishments but also for allocating liability between the cooperative and members and among members. To the extent that the cooperative is the preferred partner in the cooperative relationship, it is because the cooperative serves the administrative, organizational, and legal needs and obligations of all members regarding the digital preservation effort. Members can come and go but the cooperative must endure through technological, funding, membership, key personnel, political, institutional, legal, and other change for long-term digital preservation to occur. The rights and responsibilities section embrace that necessity in describing not only the work but also the power and rationale for accomplishing it.
Example:

The Cooperative reserves the right to remove any content that the Cooperative reasonably believes would infringe copyright or violate other local, state, federal, or international law, interfere with the MetaArchive Preservation Network, or for which the Cooperative receives a reasonable notice of potential infringement. You waive any claims against the Cooperative arising from deletions of content or other losses related to preservation of your content and any rights to physical custody of or access to any copies of content available to the Cooperative. You also acknowledge and agree that you hold full responsibility for protecting your content through alternate means beyond those copies deposited in the MetaArchive Preservation Network or resulting from your work with the Cooperative. The MetaArchive Preservation Network is not a replacement for reasonably “backing up” your content and practicing reasonable “backup” practices.

Members acknowledge that the MetaArchive, like other technology-based efforts, is not a “fail-safe” method and cannot guarantee the preservation and migration of materials. The Cooperative is one developing and important approach to digital stewardship and long-term preservation needs of the cultural and academic communities, but like other digital preservation initiatives, the Cooperative is still testing and refining strategies, techniques, and tools for resolving long-term preservation issues. The Cooperative and Members agree to exercise reasonable care and professionalism in pursuing preservation and those larger goals but make no representations or warranties otherwise regarding services and activities of the Cooperative and the Members. The Cooperative also will make good faith efforts to maintain Member access to their own materials in a reasonable and reliable manner, but events and circumstances beyond the control of the Cooperative might cause short-term or longer-term disruptions to the MetaArchive Preservation Network. The need to deploy new hardware or technology or to perform maintenance or repairs to particular caches or the larger network from time to time also may cause disruptions. You generally will receive reasonable notice of any routine and scheduled technical occurrences through email or other means requested by you for “technical notice” in your application for membership in the Cooperative. The Cooperative also will direct any “legal notice” to you at the address and to the person identified in your application materials as appropriate for receiving “legal notice.” You are responsible for providing correct contact
information to ensure that those individuals with authority to address technical or legal questions can have an opportunity to reply meaningfully to any concerns with this MetaArchive Membership Agreement or associated membership issues.

The final piece to formatting the agreement is the “signature page” The signature should include signature lines for the cooperative and members to sign and acknowledge agreement with all the terms and conditions of the membership agreement. It is also helpful to require the individual signing on behalf of the member to warrant or at least certify that the individual has sufficient authority to bind his institution and/or other members to comply with the agreement. Academic and other institutions often have highly complex rules regarding which employees or personnel can bind the institutions to the terms of any agreement for purchasing, services, or other legal obligations. Ensuring that the signed has the necessary power is essential to producing a valid and enforceable agreement.

The Makings of A Relationship – The Agreement in Practice

One central theme coursing through this analysis is that law is variable and facts change. Thus, developing a digital preservation cooperative while bounded by law is actually a series of evolving relationship requiring trust in implementation and outcome. The legal agreements and other documents for governance and technical implementation create the backdrop against which the relationships unfold. The only constant in the digital environment, including digital preservation, appears to be flexibility in adapting to change and adopting new practices as warranted or more efficient in achieving the desired goal.

The law, on the other hand, originates in an entirely different world view. Constancy and defined terms and conditions bring the certainty and predictability that makes possible achievement and success in pursuing common goals and meeting common needs. Predictability and the search for it provides the foundation for framing commercial, personal, individual, and collective effort. That predictability in law springs from a set of legal principles. The legal principles are
presumed to ensure that the project as envisioned becomes the project embodied in the agreement becomes the project as realized in the end. Pairing ever changing conditions and variables with the rigid requirements of contract law to simultaneously achieve flexibility and predictability is not only a difficulty path but it is also a seldom traveled path thus far. In some regards, the legal framework for understanding digital preservation cooperatives is as ambiguous as managing the ever evolving technological possibilities, expectations, and their corollary administration and oversight challenges.

The aim of this analysis is not completeness per se. It is instead to structure a framework for identifying and coming to grips with a seemingly infinite variety of organizational, structural, technological, and legal choices and changes taking place within the context of a fixed and enduring set of legal principles and precepts that are brought forward in time to accommodate thoroughly modern problems and issues.

The need for digital preservation arrived simultaneously with advent of digital technologies. However, the recognition that action was essential to preserve the digital record arrived somewhat later and only through the voices of few but growing bodies of concerned archivists and preservationists. The ultimate objective of this analysis is to provide a legal perspective on how to order those growing and central relationships in order to achieve widespread digital preservation while simultaneously managing legal opportunities and risks within a complex culture of mutual concern yet distrust of legal structures.

This analysis ideally brings some practical understanding of how to accomplish a result, the membership agreement, while concurrently providing a broader realization of why accomplishing it is necessary at all. Legally mature societies function in the realm of complex legal requirements. That reality is essentially a precondition for functioning in legally mature societies.

The overarching principle winding throughout the analysis is that agreements are necessary to manage risks but are more necessary to establish the trust that allows the cooperative and members to embrace and manage change while pointing to set of principles that establishes that trust in an enduring document.