# 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

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- The storage media on which the materials physically reside can fail or become out-of-date and unusable (e.g. floppy, zip, and Jaz disks)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.

[1] Hedstrom, M. & Montgomery, S. (1998). Digital preservation needs and requirements in RLG member institutions. *Research Libraries Group,* December 1998. http://www.conference-center.oclc.org/content/dam/research/activities/digpresneeds/digpres.pdf

[2] Rinehart, A. K., Prud'homme, P-A., & Huot, A. R. (2014). Overwhelmed to Action: digital preservation challenges at the under-resourced institution. *OCLC Systems & Services, 30(1), 28-42.* http://www.emeraldinsight.com/journals.htm?issn=1065-075x&volume=30&issue=1&articleid=17106334&show=html; Proffitt, M. (2011). Something's Got to Give: What Can We Stop Doing in a Time of Reduced Resources? *RBM: A Journal of Rare Books, Manuscripts, & Cultural Heritage, Fall 2011(12), 89-91.* http://rbm.acrl.org/content/12/2/89.full.pdf+html

## 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:

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

* 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.[3] 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.[4]

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."[5]

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.[6] 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.

---

[3] Daines, J. Gordon III. (2013). Module 2: Processing Digital Records and Manuscripts. *Archival Arrangement and Description*, ed. Christopher J. Prom (Chicago: Society of American Archivists, 2013)

[4] National Aeronautics and Space Administration. (2013). *The Apollo 11 Telemetry Data Recordings: A Final Report* (Final Report). Washington D.C.: U.S. Government Printing Office. Retrieved from: http://history.nasa.gov/alsj/a11/a11.html ;

[5] National Digital Stewardship Alliance. *The NDSA Levels of Digital Preservation.* Retrieved from: http://www.digitalpreservation.gov/ndsa/activities/levels.html

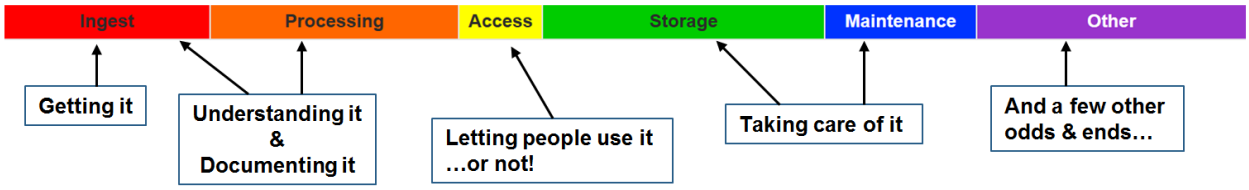
[6] Resources include the article, 'Walk This Way: Detailed Steps for Transferring Born-Digital Content from Media You Can Read In-House' by Julianna Barrera-Gomez and Ricky Erway, as well as the DP 101 page on the POWRR website.

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.[7] 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[8] and the OAIS Reference Model.[9] 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.

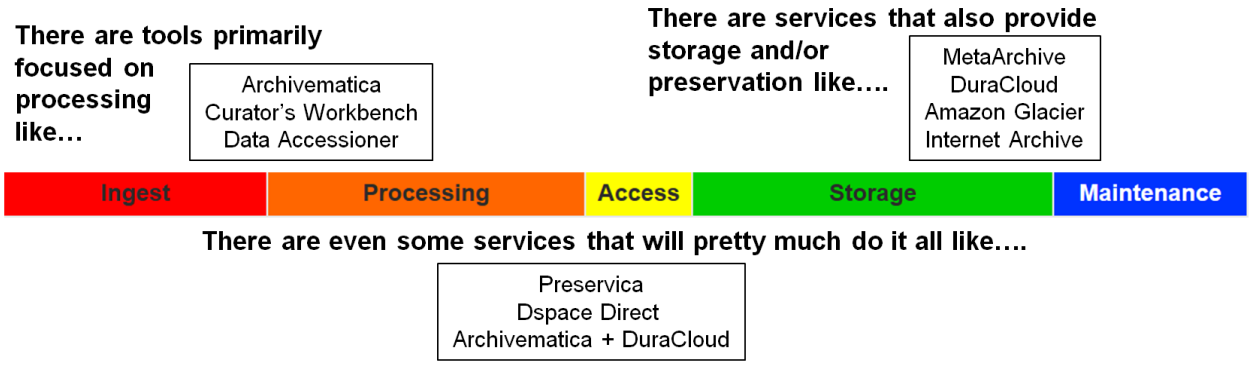
*Figure 1 – POWRR's overview of the path to digital preservation*




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.[10] 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.
[8] Digital Curation Centre. 2014. *DCC Curation Lifecycle Model.* Retrieved from: http://www.dcc.ac.uk/resources/curation-lifecycle-model
[9] Consultative Committee for Space Data System Practices. 2012. *Reference Model for Open Archival Information System (OAIS).* Retrieved from: http://public.ccsds.org/publications/archive/650x0m2.pdf
[10] Macroservices generally combine numerous microservices into a single user interface.

*Figure 2 – Different technologies often perform different functions in the life cycle (not all shown were tested in depth)*




As we began investigating some of the more robust solutions alluded to above, the POWRR team consulted with our Board of Advisors,[11] 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 | DuraCloud | MetaArchive |
| Curator's Workbench | Internet Archive | Preservica |

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.[12]

[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

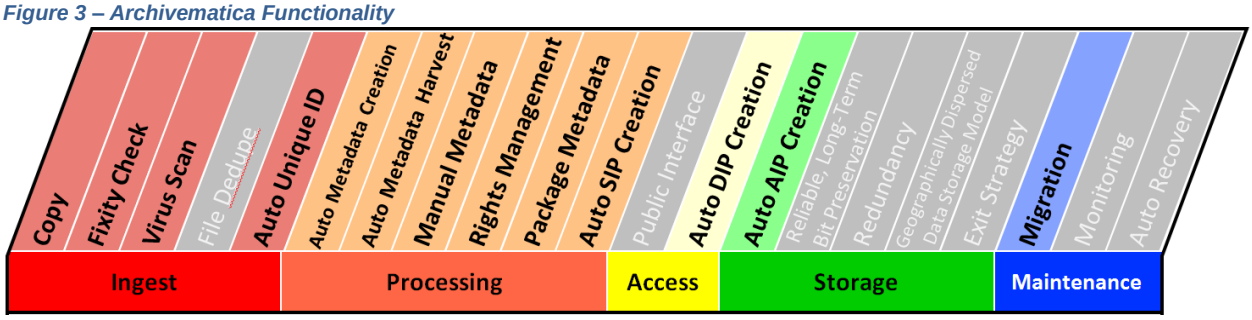
For each tool and service that underwent extensive testing by the POWRR team, we provide:

- 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[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*




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. S/he 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.[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.

[13] Consultative Committee for Space Data System Practices. 2012. *Reference Model for Open Archival Information System (OAIS).* Retrieved from: http://public.ccsds.org/publications/archive/650x0m2.pdf

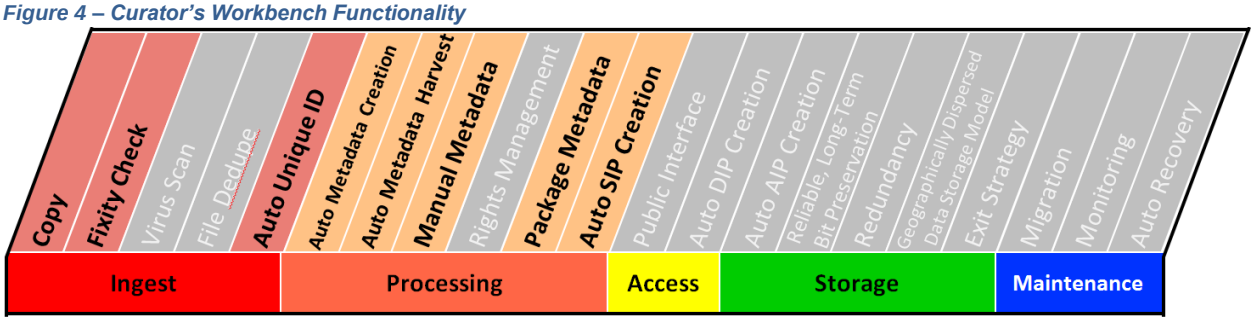
[14] We discovered these types of software do not run as quickly as typical desktop software.

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.

*Figure 4 – Curator's Workbench Functionality*




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.

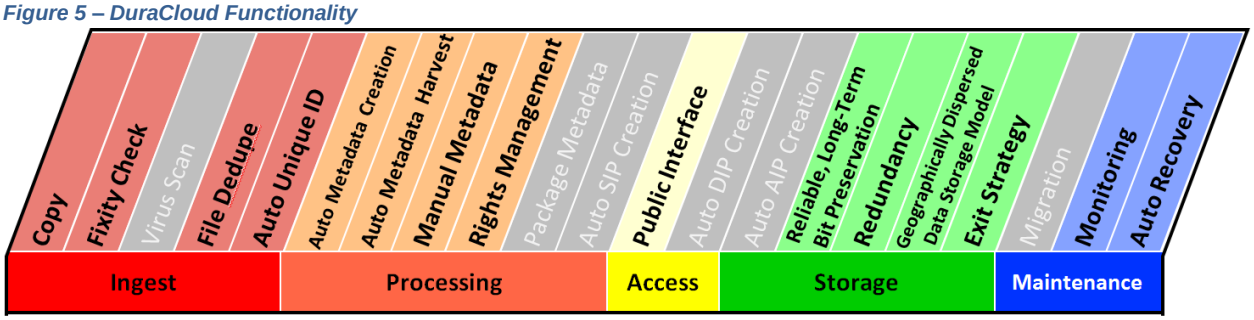
While the lead developer of this tool was responsive to inquiries, he indicated that the product is not being actively developed for a new release due to lack of current resources. There is minimal activity within the user community's Google Group. Curator's Workbench is a freely available, open source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

## DuraCloud

DuraCloud's preservation service brokers cloud-based storage[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.

[15] Cloud storage is a model in which data are stored on remote servers accessible via the Internet (aka the "Cloud")

*Figure 5 – DuraCloud Functionality*



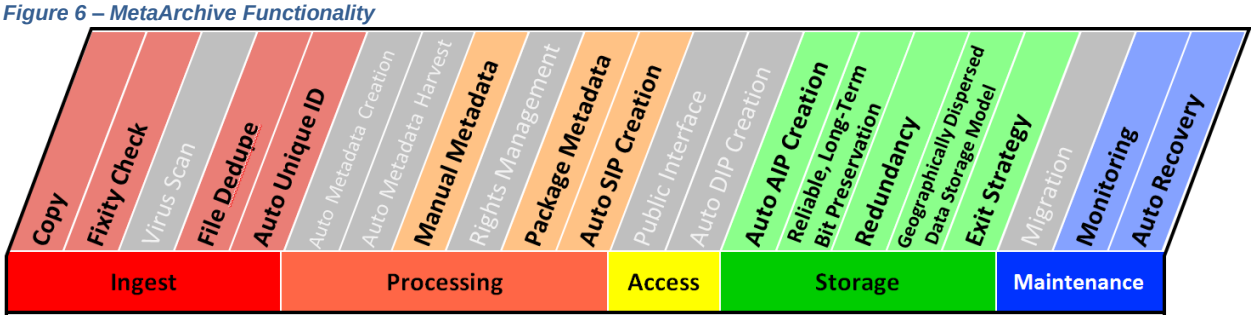

In a total of four reviews, one tester found it to be simple and easy to use, and described it as "an immediate solution to the digital preservation problem for a small institution." A second reviewer found the product to be easy to use, with complete and clear documentation available. S/he went on to 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[16] network administered by the non-profit Educopia Institute, founded in 2004, providing dark archive[17] 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*

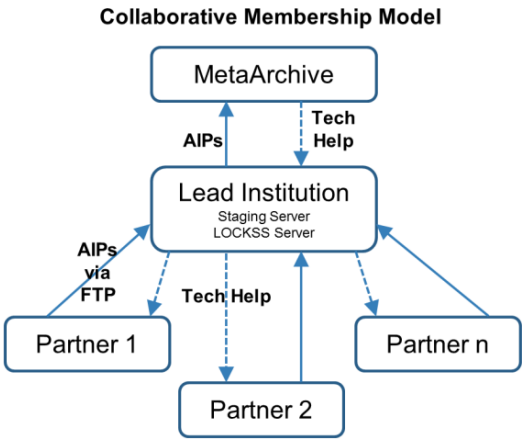



[16] LOCKSS (Lots of Copies Keep Stuff Safe) technology provides for the geographical distribution of multiple copies of digital objects within a network of cooperating institutions and is based at Stanford University Libraries.
[17] A dark archive does not grant public access and only preserves the 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.



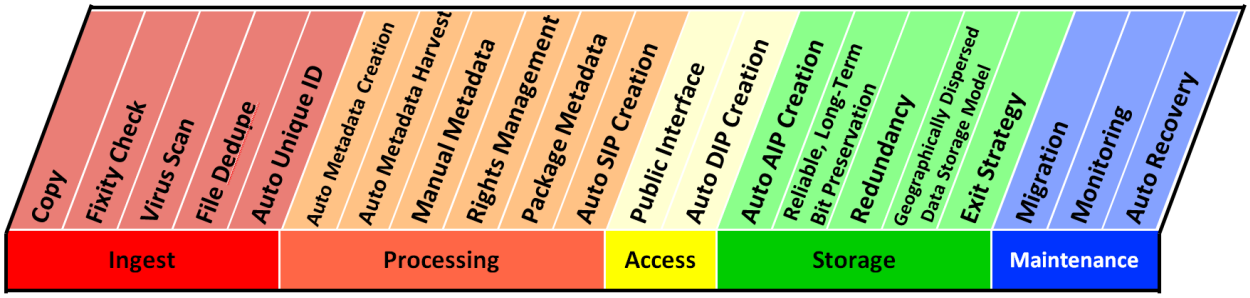

Indeed, some members indicated that being a part of the MetaArchive Cooperative was a primary reason for joining and that the actual preservation of materials was secondary.

The MetaArchive Cooperative openly provides their memberships fees and technology costs on their website: http://www.metaarchive.org/costs. 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.

*Figure 7 – Preservica Functionality*



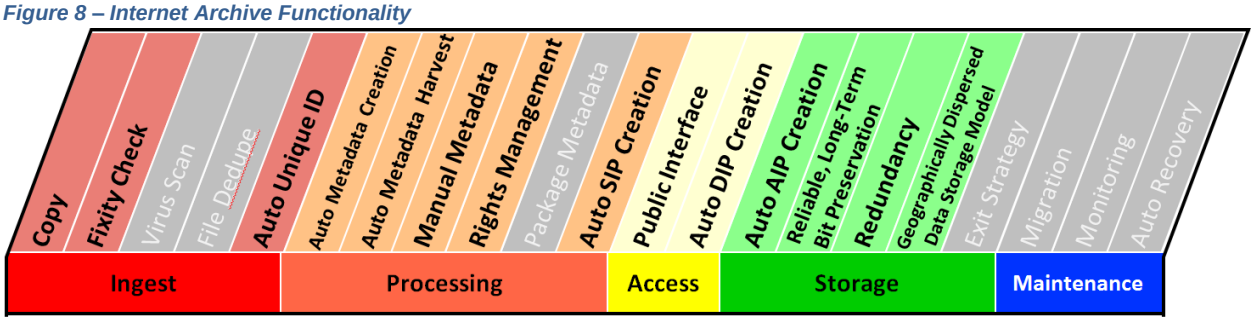

Study team members provided favorable reviews of Preservica. Reviewers found the documentation to be clear and the vendor very responsive to email questions. Although one reviewer experienced initial difficulty installing the upload tool on his/her desktop, all ultimately found Preservica to be reliable, 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]

*Figure 8 – Internet Archive Functionality*




[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.[19] 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.[20]

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 investigate 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

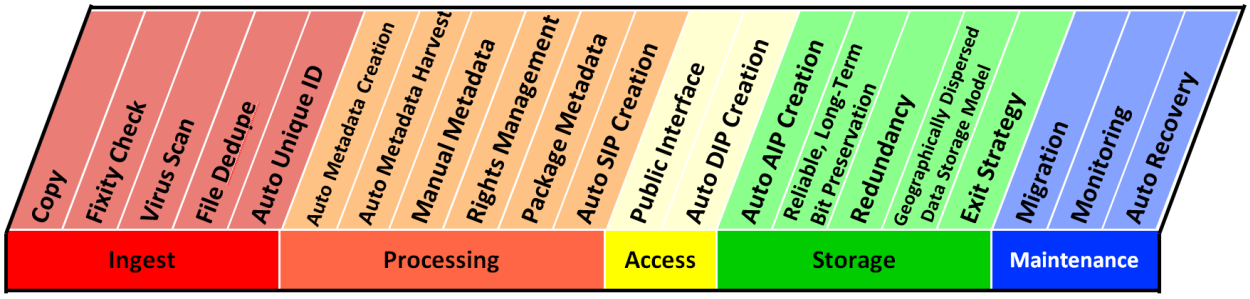
**Advisory Board**
Liz Bishoff
Steve Bromage
Martin Halbert
Jerry McDonough
Chris Prom
Amy Rudersdorf

**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 devotion 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.