Journal of Contemporary Archival Studies

Volume 2 Article 2

2015

Selecting an Electronic Records Repository Platform at the South Carolina Department of Archives and History

Brian Thomas

South Carolina Department of Archives and History, bthomas@scdah.state.sc.us

Selecting an Electronic Records Repository Platform at the South Carolina Department of Archives and History

Cover Page Footnote
Bryan Collars, digitization archivist at the SCDAH, was responsible for the hiring and supervision of the electronic records archivist for this project. Thanks is given for his guidance, patience, and latitude granted during the investigation. Thanks is also given to the rest of the SCDAH staff for their help and input during this process.

Selecting an Electronic Records Repository Platform at the South Carolina Department of Archives and History

Introduction

The South Carolina Department of Archives and History (Archives), is South Carolina's official repository for state government agency records. While it does hold a limited amount of manuscript collections, the vast majority of the Archives' collection is government records. To date, the Archives has had a policy that state agencies maintain accessibility of, and provide public access to, any electronic records (digitized or born-digital) that they have created. Given the ever more obvious need for the long-term preservation of born-digital objects, it is imperative that the Archives take the lead in handling permanent electronic records.[1]

This discussion will be about the process thus far of the establishment of an electronic records repository (hereafter referred to as the Repository), and will not discuss larger concepts such as the records continuum or appraisal and selection. It is presented from the perspective of the electronic records archivist, as one of several people responsible for the establishment of the Repository, and is not meant to take away from the contributions of other staff members. At the end of the article, a set of take-away advice from lessons learned will be given with the hope that it will help others in their own processes.

It is important to note that the Archives determined it would maintain in-state control of any records ingested into the Repository. Investment was made early on in a hardware installation with significant storage capacity. Per the Trusted Digital Repository standard, plans were made for a back-up site with a partner institution in a different area of the state for disaster recovery purposes. Given the

[1] To provide some context for the records being considered, in South Carolina records are divided into permanent and non-permanent categories. Permanent records are transferred to the Archives once their retention period has ended while non-permanent records are destroyed. The standard for permanent and non-permanent, as well as the retention period for each, is set by the Archives divisions in coordination with agencies. For some types of permanent records everything is considered archival, and in others selection is done by the Archives to ensure that records of enduring value are preserved and the remainder is destroyed.

investment in an in-state backup, a LOCKSS[2] system or cloud-based services were not considered as options.

For convenience, the steps taken in the selection process are outlined as research, policy writing, and self-evaluation/software testing. The reality of the process was messier than this, but hindsight shows this is roughly how it happened.

Step 1: Research/Literature Review

The goal of this project has been to create a Repository (tentatively named the South Carolina Electronic Records Archive) where state agencies can transfer their permanent electronic records for long-term preservation once their retention period has been met. This would take the onus off of the agency to maintain accessibility of the record, while allowing the Archives to gain control of the government's digital heritage. The Archives set the lofty goal of creating a Repository that is OAIS-compliant and meets the Trusted Digital Repository (TDR) standard.[3] It is acknowledged that meeting the TDR standard is a continuous process that may never be 100 percent achieved by any institution, but is something to aspire to.

When conducting research, the following were the guiding questions: What are the OAIS criteria/guidelines for what a repository should do? What are the TDR criteria for the model of a functioning repository? What metadata standards exist for describing, preserving, and documenting record structure, and how might they interrelate? What repository platforms exist that comply with the OAIS guidelines, TDR standard, and chosen metadata criteria?

As the state archives is not a university with a substantial set of journal subscriptions, research was limited to articles available through the American Archivist, the Australian Society of Archivists, a personal library of articles and books, and those articles freely obtainable through online sources. Even with this limitation, there were more articles than could feasibly be read while maintaining

[2] Lots of Copies Keep Stuff Safe. This typically involves a set of backup servers housed in a distributed network of archives around a vast geographic area.
[3] Consultative Committee for Space Data Systems, *Audit and Certification of Trustworthy Digital Repositories: Recommended Practice CCSDS 650.0-M-1*, September 2011.

a timeline for testing. After a thorough reading of baseline documents, a policy was adopted of monitoring listservs for useful article/blog references, regularly reviewing specific blogs, and attempting to read a published article every one to three days over the long term. This policy worked well as, even after the baseline literature was mastered, keeping up with the wealth of information in the literature required steady monitoring.

With the available resources no articles were found specifically on the process of choosing a repository. Articles most often focused on how a repository was implemented rather than the selection process itself. There was literature on components to think about when choosing a design; however, a special article reviewing just that literature would be necessary to discuss all of those resources.

During research, the following resources were the most useful:

PREMIS (Preservation Metadata Implementation Strategies) Data Dictionary for Preservation Metadata:[4] This document is a full and detailed explanation of the current preservation metadata standard.

METS (Metadata Encoding and Transmission Standard) Primer and Reference Manual:[5] This document helps explain how electronic records with multiple components can be structurally described.

Dublin Core Metadata Initiative (DCMI) metadata standards:[6] The elements in this descriptive metadata standard were common to other metadata standards investigated.

Reference Model for an Open Archival Information System (OAIS) as published by the CCSDS:[7] This document is the standard for digital repositories. It was useful in outlining the workflows and processes that will go into the Repository.

---

[4] PREMIS Editorial Committee, *PREMIS Data Dictionary for Preservation Metadata*, version 2.2, July 2012. Accessed at http://www.loc.gov/standards/premis/v2/premis-2-2.pdf.
[5] Digital Library Federation, *METS Metadata Encoding and Transmission Standard: Primer and Reference Manual*. Version 1.6, 2010. Accessed at http://www.loc.gov/standards/mets/METSPrimerRevised.pdf.
[6] Dublin Core Metadata Initiative, "DCMI Specifications." Accessed at http://dublincore.org/specifications/.

Trustworthy Repositories Audit and Certification: Criteria and Checklist:[8] This document, probably the most important in terms of guidance, provided a list of what sorts of things need to be in place and how things should work for a well-functioning digital repository.

Library of Congress (LOC) Digital Preservation Tools List:[9] This list was the reference for finding the majority of the repository software platforms that were unknown when the research began.

The Preserving Objects with Restricted Resources (POWRR) report:[10] This report lists the results of rigorous evaluations of digital repository software for archives with a small budget.

The Signal: LOC Digital Preservation Blog:[11] A place of ongoing discussion about changes taking place in the digital preservation world.

Archives and Archivists Listserv:[12] This listserv is used by many archivists in the United States to discuss problems in the field and was useful in keeping current on archives developments.

Other electronic records archivists: E-mail them, call them. Electronic records archivist are all in this problem together.

---

[7] Consultative Committee for Space Data Systems, *Reference Model for An Open Archival Information System (OAIS): Recommended Practice CCSDS 650.0-M-2*, Magenta Book, June 2010. Accessed at http://public.ccsds.org/publications/archive/650x0m2.pdf.
[8] Online Computer Library Center (OCLC), *Trustworthy Repositories Audit and Certification: Criteria and Checklist*, version 1.0, February 2007. Accessed at http://www.crl.edu/sites/default/files/attachments/pages/trac_0.pdf.
[9] Library of Congress, "NDIIP Tools Showcase." http://www.digitalpreservation.gov/tools/..
[10] Jaime Schumacher et al., "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. Accessed at http://commons.lib.niu.edu/bitstream/10843/13610/1/FromTheoryToAction_POWRR_WhitePaper.pdf
[11] Accessible at http://blogs.loc.gov/digitalpreservation/.
[12] Accessible at http://www2.archivists.org/rssfeeds.

In addition to the resources listed above, some other publications provided good background into key concepts for a functional repository and should be mentioned. In 2000 Patricia Galloway produced a report that discusses how a state archive might go about collecting records from agencies, what metadata should be addressed, and what legislation or policies might be necessary to ensure archives are able to collect records.[13] For state agencies, this is *highly recommended* reading. In 2002 LeFurgy discussed how a repository should define exactly what services it will provide to records depositors.[14] The InterPares Project, a continuing source of archival research into electronic records, produced a grant report in 2002 with workflows and ideas on how to maintain evidence of authenticity of electronic records.[15] In 2004 the Electronic Resource Preservation and Access Network (ERPA) set guidelines on ingest strategies.[16] In 2008 Hockx-Yu and Gareth produced a report summarizing a workshop on what significant properties of a digital object may need the most consideration in preservation.[17] In October of 2014 Lavoie published an easy to follow introductory guide to the Open Archival Information System (OAIS) reference model.[18]

Step 2: Policy writing

Policy writing came before testing. Trusted Digital Repository (TDR) policies were created before investigating actual software because (1) it was an easier task and (2) it set the standards by which to judge the software options. Using research with the literature, combined with known time and duty limitations of agency records managers, policies were developed that worked toward the TDR standards while providing a limited burden on records managers. Policy writing fell into a few general categories: (1) general policies on handling situations, procedures,

[13] Patricia Galloway, *Mississippi Electronic Records Initiative: A Case study in state government electronic records Final Report*, Mississippi Department of Archives and History, May 2000.
[14] William G. LeFurgy, "Levels of Service for Digital Repositories," D-Lib Magazine 8, no. 5 (May 2002). http://www.dlib.org/dlib/may02/lefurgy/05lefurgy.html.
[15] US-InterPARES Project, *Findings on the Preservation of Authentic Electronic Records: Final Report to the National Historical Publications and Records Commission (Grants #99-073 and #2001-005)*, Sept. 2002.
[16] Electronic Resource Preservation and Access Network, *Erpa Guidance: Ingest Strategy*, September 2004. http://www.erpanet.org/guidance/docs/ERPANETIngestTool.pdf.
[17] Helen Hockx-Yu and Gareth Knight, "What to Preserve? Significant Properties of Digital Objects," *International Journal of Digital Curation* 3, no. 1 (2008): 141–53.
[18] Brian Lavoie, *The Open Archive Information System (OAIS) Reference Model Introductory Guide*, 2nd ed. Digital Preservation Coalition, October 2014. http://dx.doi.org/10.7207/twr14-02.

and workflows; (2) application of wider standards to the local instance of the Repository; and (3) metadata standards to be used for description, structure, and preservation. This last aspect in particular was important because it was the basis for judging whether a repository software platform would aggregate the metadata identified as critical. If serious investigation of a repository system had been done before this step the Archives would have run the risk of deciding its metadata standards based on what the software options allowed for, rather than what was really needed.

After much deliberation, the Dublin Core-based description standard was chosen because of its simplicity and integration with other metadata standards. Requiring Dublin Core metadata from agencies did not seem to be over-burdensome on agency records managers, nor did it require they have special knowledge of a subject. Also, using Dublin Core maximized the possibility for expansion and adaptation to other standards.

Just as a granular specialized metadata schema was determined to be too burdensome, descriptive data at the individual record level was deemed inappropriate. To maintain the context of a record set and limit the time necessary to vet records before ingest, an "intellectual unit" (IU) was selected as the appropriate level of description. While technically this could be a single digital object, it was meant to serve as a container description for records grouped around a specific theme or event. For example, if an agency groups its electronic records around a case, the case file is the IU. Based on the existing accession transfer form and series description fields, a draft metadata schema was set in place for repository software evaluation.

Step 3: Self-evaluation and software testing

3a: Self-evaluation

Self-evaluation of the Archives' capacities went hand-in-hand with software identification and testing. It was an evolving process that is only separated from testing here because it is important to highlight some insights from the process. As inconvenient as it is to ask what a facility is capable of supporting, the reality

is that the resources of every archive are different. Evaluation began as a basic assessment that evolved into the question of feasibility of implementation.

Upon reflection, the following were guiding questions in software evaluation:
Should the inquiry stick to open-source software or is proprietary software an option?
How much of a deciding factor should dissemination/access capability play?
What is the infrastructure hardware available once a software platform is chosen?
What kind of human technical support will be available to maintain the software and hardware system?
What is the threshold for removing software from consideration?
How much stitching together of tools is permissible before an implementation is considered impractical?
How are maintenance/subscription fees measured?

Of the questions listed above, the primary driver was human resources and support/maintenance. The Archives was on a recovery track after cutbacks due to the 2008 recession. At the time of this writing the Archives has only two staff IT personnel stretched over several divisions (Archives and Records Management, and Historic Preservation) and two buildings. Candid reflection on available staff time showed that software would need to be maintained by the electronic records archivist or by purchase of a maintenance agreement/subscription. As an agency funded by taxpayers, maintenance costs were not considered lightly.

A significant but secondary component of the repository evaluation question was access. Without public access to preserved materials (or their surrogate), it is as if preservation never took place. Any solution would need an access component or have the possibility of integration with an access tool. The combination of both preservation and access software needed to result in a feasible annual cost. A contextualizing, hierarchical interface such as Dspace was considered ideal. The access component was listed as secondary due to the limited options for repository software that supports this need.

By the time of testing, three tiers of repository options were considered viable:

1. An integrated repository software platform that includes automated digital repository services with an integrated access tool.

2. An automated/semi-automated digital repository software platform, with the possibility of manual connection of dissemination packages for public access, or access mediated through archives staff.
3. Manual preservation with a variety of tools leading up to overall digital preservation with or without access.

Tier three was considered a last resort. Since there are a multitude of open-source tools available, a benefit is that a broken tool could be replaced by one with similar functionality. In addition, maintenance fees for individual small tools is unlikely. While using open-source tools collected together does appear to address the Archives' IT limitations, it does not take into consideration the processing time for the archivists. An archive with minimal electronic records to preserve could take time for manual preservation work, but it is not easily scalable to a state archive level. As the plan is to get all state agencies to begin submitting records, the system needed to be designed with a large ingest capacity in mind.

3b: Testing

Even though a good-faith effort was made to track down and test available repository software options, several tools have probably been unintentionally overlooked. Repository software seems to be stabilizing into a limited list of well-developed options, but it is still somewhat of a wild west with many tools available to address customized needs. Thus, this list of tools evaluated should be considered with the knowledge that it may not be all-inclusive.

In an effort to determine feasibility of maintenance by the Archives and functional capacity of the software, all options were installed by the electronic records archivist on a testing computer using virtual machine environments. Virtual machines were a very useful tool as they allowed multiple attempts at installations when something went awry. It allowed scaling up or down the size of an installation and prevented conflicts between preservation platforms.[19] Also, by running multiple virtual machines, side-by-side comparison between platforms was possible.

[19] Many of the software platforms use http://localhost for web-browser access to the system. Multiple platforms on the same machine led to irresolvable issues.

In the VMware Player program,[20] the first step was installing the basic operating system, almost always Linux Ubuntu 12.04,[21] and completing the necessary updates after installation. After the operating system installation, the preservation software platform documentation was followed until the platform installation was successful. If a failure occurred and was unresolvable, an attempt would be made to identify the point of error, then the virtual machine was deleted and the lesson learned was incorporated into the next attempt at an installation. It was not uncommon to have three or more attempts at an installation before being successful.

In most cases, evaluation was stopped as soon as it became apparent that the platform would not meet the Archives' needs.[22] Thus, a functional grid that represents every aspect of a tool is not possible. Instead, the pros and cons are explained as they became apparent during the evaluation process.

DAITSS.[23]

This software platform, available from the University of Florida, is free and open-source. It is strictly a preservation platform.

Pro:
Free and open-source.
Software neutral with bit-level preservation.
Creates preservation and structural metadata.
Very stable without need for any subscription service.

[20] VMware Player was chosen over VirtualBox, the other major virtualization program, due to personal preference. Getting VirtualBox to access some computer components such as the 3.5" floppy drive was difficult, but that is likely due to user error. VMware Player is available for download from the VMware website at http://www.vmware.com/player/.
[21] Available at http://www.ubuntu.com/.
[22] The "hit by a bus" rule applied in testing. This rule dictates that if an employee is "hit by a bus" and is suddenly unavailable to explain things, then the employee should have made sufficient resources available for other employees to know where things stand and how to move forward. Even though it might be possible to develop the necessary skills for keeping the repository running, if it became obvious that too much specialized knowledge was necessary for maintenance by the rest of the staff then a maintenance agreement was required from a vendor.
[23] Available at http://daitss.fcla.edu/.

Con:
No public dissemination piece.
No state agency submission piece.
No automated normalization[24] to other formats.
Does not provide for creation of derivative files.

Assessment: As DAITSS did not meet the needs for normalization and access, it was not considered the right tool for the Archives. It is worth noting that the Premis In Mets Toolbox[25] used as a part of DAITSS is an excellent tool for the creation of well-formed Premis metadata when doing manual preservation work. Any archive considering use of manual preservation methods should think about implementation of that tool.

Digital Preservation Software Platform (DPSP).[26]

This software platform, originating from the National Archives of Australia, is a composite of several other tools that taken together provide a repository service. The tools are Xena, Digital Preservation Recorder, Checksum Checker, and Manifest Maker.

Pro:
Stable and does not appear to need any maintenance.
Does normalization of certain file types to open-standards formats.
Saves to its own native container format to prevent tampering by outside users.
Provides records with unique identifiers.
Runs checksums and virus scans.
Provides bit-level preservation.

Con:
No public dissemination piece.
Output not structured for archives staff to facilitate access to the materials.
No state agency submission piece.

[24] Loosely defined, normalization is the process of taking a type of file format and either ensuring its compliance with format standards or transforming it to a pre-established standardized format.
[25] Available at http://pim.fcla.edu/.
[26] Available at http://dpsp.sourceforge.net/.

Items submitted need manual cataloging outside of the system to keep track of content.

Assessment: DPSP is a stable and wonderful tool. However, manual submission by archives staff with the need to create associated descriptive metadata outside of the system was a deal-breaker. This presented a significant potential for failure if the outside database was not properly maintained, or somehow unique IDs became disassociated with catalog records. The lack of easy access to the public was also problematic.

FEDORA.[27]

Fedora is a powerful repository software platform that is format neutral and allows great flexibility in how it is applied. It is a stand-alone tool, and is very easy to install with some background knowledge in Apache Tomcat and databases.

Pro:
Free and open-source.
Very widely adopted in the university environment, as well as other types of institutions.
Very flexible in application and interaction with other tools.
Supports versioning of objects.
Stable.

Con:
No normalization.
Requires plug-ins for ingest and preservation activities.
Direct public access is possible, but not great.
Manual ingest or command line ingest is too burdensome for agency direct submission.
Metadata ingested with materials; not a completable form when ingesting.

[27] Available at http://www.duraspace.org.

Assessment: Fedora is a good concept, but it is designed to be part of an ecosystem where access, ingest, and normalization are done by other tools. Its support of versioning means normalization can be done, but not by the repository platform on its own. Ingest of metadata normally comes through an XML document so it would be challenging to get the average records manager to provide metadata. On its own, it is too difficult to use and would need to be integrated with other tools. There are integrations available.

Dspace.[28]
Dspace is a open-source and free repository software program widely adopted by universities for submission and preservation. Some archives use it as a dissemination tool as the public interface is strongly hierarchical, just like collections tend to be.

Pro:
Virus scan.
Fixity checks.
Faceted browsing.
Reports on format types.
Remote submission by agency staff.
High level of descriptive metadata possible.
Contextualization through nested hierarchies.
Ability to limit access/submission based on credentials.

Con:
Does not provide normalization.
Somewhat clunky submission process.
Does not provide format migration options.
Dissemination copy is the same format as submitted copy unless intervention is done.

Assessment: From prior knowledge of Dspace, it seemed like it would be the right fit. However, after many attempts an installation still was not running properly, which meant that external support would be necessary. There are several service

[28] Available from http://www.duraspace.org.

providers that can help with installation and maintenance for a reasonable fee. Dspace provides fixity checks, virus scan, and strong metadata; and it is unparalleled for providing contextual relationships. For preservation purposes, it was not sufficient since most normalization would need to take place outside of the repository with manual reingest of records.

Hydra.[29]
Hydra is a Fedora integration framework with a flexible application structure. The idea of Hydra is to have multiple heads of access and ingest linking back to one Fedora repository using specific standards.

Assessment: Ingest and access tools are separate plug-ins. The framework was too complex to assemble in-house with current staff expertise. This software was not given significant consideration because there were too many moving parts for testing and the potential for breakdown in implementation was deemed too significant with limited staff.

Islandora[30]
Islandora is a Fedora integration that uses Drupal CMS as its front-end for access as well as ingest.

Pro:
User friendly with a Drupal interface.
Relies on the Fedora repository software platform, which is very stable.
Inserts the metadata generated in Islandora into the Fedora Repository.

Con:
Digital objects rejected if its format extension (jpg, png, etc.) is not registered by administrator.
Object types need to be declared when ingesting, leading to potential bottlenecks.
Batch submission was very tricky.
No virus scan or quarantine functionality built in.
No normalization of formats.
Installation of dependencies was problematic.

[29] Information available at http://projecthydra.org/.
[30] Available at: http://islandora.ca/.

Complex digital objects were difficult to create.

Assessment: Do not let the list of cons suggest that Islandora is a bad tool. Islandora is an amazing tool with a wonderful public interface. Drupal was researched independently of Islandora as a potential unified public access piece for the collections. It was found that while Drupal has great potential, significant problems would need to be handled out of house. It took several configuration attempts to get the connection going between Drupal and Fedora. In the process of installing the software dependencies, enough errors occurred that help would be needed in installation to get the full functionality. There are companies that do Drupal support, but no companies could be found that did support for an Islandora/Drupal instance as a normal part of their business. If there is high level Drupal expertise in-house, then this software is worth investigating. However, for the Archives support would be needed for installation and maintenance. As it did not also include easy normalization and easy ingest functionality, it was not an avenue that made sense.

Archivematica/AtoM.[31]
Archivematica is a free and open-source tool developed and maintained by Artefactual. It does not have an internal dissemination piece, but interacts with other software (Archivist Toolkit, for example) to automate uploads of dissemination copies of electronic records. AtoM is Artefactual's archives catalog software and Archivematica's de facto dissemination platform.

Pro:
Open-source software that contributes to the larger community.
Uses open-source tools to normalize into preservation formats.
Net cost significantly less than proprietary options in the near-term.
Easy to customize workflow for digital preservation.
Easy to use.

Con:
Self-installation is tricky but possible.

[31] Available from the software developer Artefactual (http://artefactual.ca). The Archivematica platform is available at: https://www.archivematica.org/wiki/Main_Page. The dissemination platform AtoM is available at: https://www.accesstomemory.org/en/.

Implementation would require a maintenance agreement for at least a limited period of time.
Open-source tools only; no independent development for normalization.
No direct public interface; uploads to other interfaces.
Internal staff navigation is through search or item-level browse; hierarchical navigation is not possible.
Reingest of a whole Archival Information Package is necessary to add a derivative copy of a record to an existing item in the repository.
In testing, unable to connect AtoM and Archivematica.
Upload options are Archivists' Toolkit, AtoM, and ContentDM, none of which are used by the Archives.

Assessment: As with Islandora, do not let the list of cons be the deciding factor. Archivematica is an amazing product. Of the open-source tools available on the market, it is by far the best for preservation workflows on the repository side. Getting Archivematica installed was initially quite difficult and necessitated troubleshooting a quirky permissions problem with a critical Archivematica file in order to get full functionality. Once installed, the system works beautifully, smoothly, and with ease. If an open-source tool were used, this would be the best choice for the Archives. However, it is based entirely on open-source tools and extra functionality is based on special projects that cost additional money. It has active development and support, but is limited in some ways by its philosophy of open-source tools.

The problem from the Archives' perspective was that the native dissemination piece suffered from the need to create a new entry in the AtoM system for every IU ingested as a whole. An entire CD-ROM could be uploaded, but in order to show that as an IU, a sub-series (or file, folder, etc.) needed to be created. Otherwise, the link would be to the series level and the intellectual context could be lost. Also, additions of preservation or access derivative objects to an Archival Information Package in the repository would be through a reingest of the whole package or creation of an artificial collection. Outside preservation work would need to occur before ingest, not afterwards.

Preservica.

Pro:
Development requests are queued; does not need special project payments.
Developed features contribute to future release enhancements.
Regular iterations with new functionality added.
Customization of workflows.
Integrated public access piece that automatically pushes records to external users.
Portable submission utility for installation on agency records manager computers.
Ability to manually add a derivative file to an existing Archival Information Package.
Preservation copies/derivative copies can be made at any time after initial ingest.
Internal user interface allows after-ingest metadata additions and hierarchical browsing to a collection.
Graphical reports.

Con:
Significant cost.
Proprietary development; does not contribute to global community.
Public interface a bit difficult to navigate.
Specialized metadata schema with embedded additional metadata.

Assessment: Preservica's main detractions were the cost and its proprietary nature. A local installation comes at a significant fee.[32] To get the benefits of further developments, there is a continuing maintenance fee. While it does share its developments with the internal user community, advancements are less frequently shared to the field as a whole. Despite these drawbacks, in consideration of the positive qualities of the overall platform, its cost and proprietary development structure were considered acceptable.

Testing was done on a cloud version of the platform for a fee, so it is unclear how difficult everyday maintenance would be. However, maintenance fees do include software support. While the cost was prohibitive, the platform was determined to

[32] A cloud edition is available for significantly less. However, it has some additional limits to functionality. As hardware infrastructure had already been invested in, a cloud edition was not feasible.

be both scalable for state government and user-friendly enough that changes in staff expertise would not significantly impact its functionality.

Preservica won out over Archivematica in its ability to facilitate public access. While the public interface is somewhat difficult to navigate, for the scale and variety of records expected by the Archives, it worked better than the alternatives. The ability to push records to and from public access as necessary allows for easy roll-outs of large record sets once they are ready for the public. In addition, browsing through hierarchies on the internal user interface created a system that can mimic the Archives' series/record group structures. The platform was the single source for these types of access functionality.

Advice

Some advice to those investigating digital repository options:

1. Do not panic

There is quite a bit of literature out there about electronic records preservation, more than most can master without years of time in the field. Carry out due diligence in research, but accept limitations in time for research and triage to sources that seem most applicable.

If there is not a lot of technical support or funding, be sure to read the POWRR report to get a feel for the most cost-effective common options. Bear in mind that these options may not be cheap but could be cheaper than hiring additional IT staff. Due diligence may still be needed to establish the viability at the institution, so this should not be the final source for making a decision.

Reread those manuals about metadata schemas in light of the intended project. Thinking of a practical application will produce new insights and provide a functional list of the basic metadata required.

2. Know repository standards.

The TDR guidelines and OAIS guidelines give good rules to follow; read them. Reread them. Keep the guidelines in mind when investigating software; it will help with the process.

3. Establish functional requirements.

If there is an expectation of a limited volume of materials and archivist mediated ingest is preferred, the DAITSS or DPSP repositories might be a good fit. If a vastly distributed remote submission capacity with users is required and those users can take the appropriate time to submit materials with metadata, Fedora or Dspace could be the best options. The Archives was constrained by potential submitters with limited time to create metadata, the need for direct public access, and IT staffing. Others might not be.

4. Set up a testing environment.

Nothing beats a hands-on approach to software investigation. Ideally, there will be a virtual machine capacity so that, for example, the need to install Linux does not require overwriting a Windows-based operating system. With virtual machines a fresh install is *much* easier, so an irrecoverable error does not ruin an entire computer.

5. Talk to people.

Coming into contact with other electronic records archivists, it became abundantly clear that most people are still struggling with the same issues. In overcoming their own obstacles, they might have chosen tactics that others might not have thought of.

Summary

Due diligence with investigation and testing of repository software options was necessary for the establishment of the South Carolina Department of Archives and History's electronic records repository. Testing was a laborious process. It

involved many pitfalls, problems, frustrations, and the occasional joy of success. It was rewarding to learn from each failure how to do things better the next time when testing a type of program. Lack of success was often the indicator that was required in determining the viability of a platform.

As of the time of this writing, the Archives has chosen Preservica as its final solution. Although a public agency's choice of proprietary software is counter-intuitive, the platform's ability to manage large record sets and push those records to the public facilitates the greater open-government goal of public access. Also, its support structure provides a level of insurance against the possibility of changes in staff expertise or cutbacks, which could otherwise result in disrupted service. Installation is currently pending acquisition, hardware installation, and a disaster recovery site being established. Once these elements are completed, the Archives should be on the right track to reaching a Trusted Digital Repository status.

Post-implementation the plan is to actively collaborate with records managers to gain traction with state agencies and begin accessioning born-digital records. This is already happening on its own as agencies are hearing about the Archives' work to establish a repository. A big hurdle will be determining the most agency-friendly means of records transfer.

References

Consultative Committee for Space Data Systems. *Audit and Certification of Trustworthy Digital Repositories: Recommended Practice CCSDS 652.0-M-1*. September 2011. http://public.ccsds.org/publications/archive/652x0m1.pdf

Consultative Committee for Space Data Systems. *Reference Model for an Open Archival Information System (OAIS): Recommended Practice CCSDS 650.0-M-2*. Magenta Book. June 2010. Washington, DC: CCSDS Secretariat, 2012. http://public.ccsds.org/publications/archive/650x0m2.pdf.

Digital Library Federation. *METS Metadata Encoding and Transmission Standard: Primer and Reference Manual*. Version 1.6. 2010. http://www.loc.gov/standards/mets/METSPrimerRevised.pdf

Dublin Core Metadata Initiative. "DCMI Specifications." http://dublincore.org/specifications/.

Electronic Resource Preservation and Access Network. *Erpa Guidance: Ingest Strategy*. September 2004. http://www.erpanet.org/guidance/docs/ERPANETIngestTool.pdf.

Galloway, Patricia. *Mississippi Electronic Records Initiative: A Case Study in State Government Electronic Records Final Report*. Mississippi Department of Archives and History, May 2000.

Hockx-Yu, Helen, and Gareth Knight. "What to Preserve? Significant Properties of Digital Objects." *International Journal of Digital Curation* 3, no. 1 (2008): 141–53.

Lavoie, Brian. *The Open Archival Information System (OAIS) Reference Model: Introductory Guide*. 2nd ed. Digital Preservation Coalition. October 2014. http://dx.doi.org/10.7207/twr14-02.

LeFurgy, William G. "Levels of Service for Digital Repositories." *D-Lib Magazine* 8, no. 5 (May 2002). http://www.dlib.org/dlib/may02/lefurgy/05lefurgy.html.

Library of Congress. "NDIIP Tools Showcase." http://www.digitalpreservation.gov/tools/.

Online Computer Library Center (OCLC). *Trustworthy Repositories Audit and Certification: Criteria and Checklist*. Version 1.0. February 2007. http://www.crl.edu/sites/default/files/attachments/pages/trac_0.pdf.

PREMIS Editorial Committee. *PREMIS Data Dictionary for Preservation Metadata*. Version 2.2. July 2012. http://www.loc.gov/standards/premis/v2/premis-2-2.pdf.

Schumacher, Jaime et al. "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. http://commons.lib.niu.edu/bitstream/10843/13610/1/FromTheoryToAction_POWRR_WhitePaper.pdf.

US-InterPARES Project. *Findings on the Preservation of Authentic Electronic Records: Final Report to the National Historical Publications and Records Commission (Grants #99-073 and #2001-005)*. September 2002.