4-12-2016

Proposal for a Digital Archives Program at the Dr. Joann Rayfield Archives

Eric Willey
Illinois State University, emwille@ilstu.edu

April Karlene Anderson
Illinois State University, aander2@ilstu.edu

Ross Griffiths
Worcester State University, rgriffiths@worcester.edu

Dallas Long
Illinois State University, dlong@ilstu.edu

PROPOSAL FOR A DIGITAL ARCHIVES PROGRAM AT THE DR. JOANN RAYFIELD ARCHIVES

by

April Karlene Anderson, Ross Griffiths, and Eric Willey

Date: April 12, 2016

PREAMBLE

Dr. Jo Ann Rayfield University Archives collects, organizes, and describes millions of documents related to ISU's history and makes them available for researchers both on and off campus. This documentary output has accrued since the university's founding in 1857, but its format has changed dramatically over the past 30 years. Since the 1980s, digital files have replaced paper as the format of record for an increasingly broad number of functions across all administrative and disciplinary areas. Today, ISU's history is no longer recorded exclusively in manuscripts, books, and photographs but also in digital files created in a variety of different formats, stored to different kinds of media, and dependent on a variety of software, hardware, and operating systems to function. As with analog records, these digital files are vulnerable to loss through disposal, decay, and disaster but their dependency on complex and interrelated technologies requires new archival policies, procedures, and practices to manage them.

As the official repository for ISU's documentary history the Dr. Jo Ann Rayfield Archives finds itself lacking the critical infrastructure necessary to preserve these born digital and digitized records. The Archives proposes the following plan to create the infrastructure necessary to collect, preserve, and provide access to the digital content that has been and will continue to be created by the institution. The plan will define the digital objects Archives will ingest, how the objects will be managed, and how the digital archives will integrate with the physical holdings; furthermore, three software and hardware solutions will be provided to offer a range of methods to implement that plan.

FORMAT AND TECHNOLOGY OBSOLESCENCE

A central problem for digital archives is technical obsolescence. Because digital data is machine-readable, accessing a digital object requires the use of appropriate hardware, software and operating

systems. The rate of technological change in recent decades has rendered many older systems obsolete and the files created in them inaccessible.

As manufacturers continually introduce and update products, files created in older generations of software or stored on obsolete media are often unreadable on current computers. Popular computer hardware also changes rapidly with newer desktop computers equipped with storage devices and peripheral connections that are often incompatible with older technology. Of the dozens of types of digital storage media used over the past 50 years, MIT has identified 32 different types that are either totally obsolete or at risk for imminent obsolescence, many of which were ubiquitous as recently as 15 years ago. [1]

To address this problem, archivists may assemble a collection of computers equipped with older software and hardware systems on which they can open files and reformat them to current specifications. Maintaining an array of recently outdated systems and software such as this is a feasible solution for most institutions, but accessing much older or more complicated technologies may present significant challenges and require the use of a vendor. In addition, developers have created some open source programs that emulate the functionality of obsolete older software and which may be used to open various obsolete files.

Digital Preservation

The success of any digital archives program depends on the larger technical infrastructure in which it functions. This infrastructure includes both hardware and software products, but it also includes the systems in which digital archives are stored. Long term digital preservation depends on two separate but interrelated functions: file redundancy and file integrity. In order to protect digital collections against the catastrophic failure of a storage system, all digital files in a collection must be duplicated in a separate and unrelated storage system. These redundant systems may include a local

[1] ibid. Chamber of Horrors: Obsolete and Endangered Media
http://www.dpworkshop.org/dpm-eng/oldmedia/chamber.html

network in combination with a cloud service, or separate cloud services, or separate networks. The files stored in these systems must also be periodically checked for integrity to ensure they have not been tampered with, deleted, or corrupted. Because the complexity and volume of digital collections precludes a simple visual inspection of data, a technical function called a checksum is used. A checksum produces a unique numerical "fingerprint" of each file or digital collection which is then repeated at regular intervals. Any changes in the checksum results indicate changes or corruption of data. The Archivist will then replace the problem data with a copy of valid, uncorrupted data from one of the redundant digital storage systems.

The Situation at ISU

At ISU, much of the data produced since the mid-1980s was created on desktop computers using various technologies many of which are now obsolete. Hardware and applications for word processing, spreadsheets, and databases include obsolete versions of the Microsoft Office Suite as well as archaic or defunct products such as WordPerfect, Lotus 123, dBase, AppleWorks, and Starburst. Digital storage media includes optical media such as CDs and DVDs, various kinds of external drives, memory cards from digital cameras, as well as different sizes of floppy disks for which there are no drives or software currently available. Magnetic tape reels from mainframe computers are not currently part of the collection, but are in the possession of some campus departments and may eventually be accessioned by the Archives.

In the short term, the Archives should assemble a collection of available older versions of desktop software reflecting the most typical file types found in the collection as well as a list of vendors that might be able to service unavailable, obsolete files and formats. In the longer term, 3" and 8" floppy drives and the machines and software necessary to run them should be obtained.

Special Collections at Milner Library does not currently receive significant amounts of born digital or digitized files from donors. The small amount of digital material which is accessioned is often

on optical media and part of collections which primarily consist of materials in physical formats. The majority of digital files that are accessioned are from Gamma Phi Circus and consist of digital photos and video of their annual show. These files are not frequently accessed. Therefore, at this time, backup of the limited amount of digital files acquired on physical media to the digital archives as a check against data corruption (most likely due to degradation of the physical media) is adequate to meet the needs of the Special Collections Department. It is not anticipated that frequent access will be required such that Special Collections would require a terminal and software for on-site viewing.

Special Collections does generate a large number of digital files created through digitization of existing manuscript holdings (primarily for display in CONTENTdm or as working or access copies). As these are digital surrogates and not in fact archival materials in and of themselves, existing backup measures (placement on the institutional R: and H: drives, and procedures followed by the digitization unit for CONTENTdm material) are deemed adequate for these materials. While the network preservation capacity of ISU is an ongoing concern, the original physical copies of these objects have been retained and could be rescanned if the data were lost.

This issue may be revisited in the future should the amount of digital materials acquired by Special Collections increase. In this case, it would likely be necessary to have the ability to ingest and particularly to provide access on site (in Milner Library, preferably in the Special Collections reading room). This would require appropriate hardware, software, and training of staff. There is no timeline for implementation of this option, as it will be driven by donor and patron demand.

At present, the capacity of the ISU network to provide digital preservation is unclear. Published policies and practices related to general data storage, redundancy of data, data recovery procedures, system backup, emergency plans, and security policies are not available to Archives or Library personnel. Until the time when the library and archives can view clear, documented, and audited practices that demonstrate the ISU network's ability to meet appropriate standards for digital preservation, the library

should obtain an appropriate vendor-based digital storage solution for at least one full instance of all digital archives.

DIGITAL ARCHIVES WORKFLOW AND TIER LEVELS

Listed below is a proposed workflow for a digital archives repository and tier levels for implementation of the repository.

Workflow

The following is a proposed workflow for the processing of digital objects and making them accessible for patron use. The workflow is for a Tier I digital archive (see tier listing below) though the basic tenants of the work can be applied at all tier levels.

1. *Getting the materials*

   Similar to physical collections, digital collections will go through an established set of procedures for accessioning new materials into the University Archives collections. Digital objects will be received by the University Archives in a variety of ways including via email attachments, sent through a File Transfer Protocal (FTP) system like SendTo or downloaded from a cloud service, and physically given to staff on portable media (optical media disks, flash drives, portable hard drives). Vital information including collection title, date of acquisition, file size, file types, and other fields will be entered into a dedicated inventory XML file. The new acquisition will also be given a unique identifying number for tracking purposes.

Location Information	Donor Name	Title	Creators	Creator Authority	Archival History	Scope and Content	Appraisal, Destruction, & Scheduling	Physical Conditio
Paper - North Storage Stacks; Digital - MLB_Archives\Archives\DigArch\Farlee_Lloyd	Tim Fredstrom - Dept. of Music	Lloyd Farlee Papers	Farlee, Lloyd	Local	The materials were transferred to the Archives by Tim Fredstrom. An alum of the program scanned all of the original documents to a PDF format sometime in 2011. The originals were processed and made available for research, Fall 2012. The digital materials were held on a portable disk drive until they were transferred to the Archives	Collection was organized into two series: Arrangements (I) and Personal Materials (II). Collection is mostly handwritten scores.	No materials were excluded from the collection upon its acquisition to the Archives.	Physcial materials were preserved in acid free fo boxes at time of process Scanning equipment use files unknown; files not Likely no OCR.
Paper - North Storage Stacks (F/7/1-6 & F/8/2-4); Digital - MLB_Archives\Archives\DigArch\JGAPE	Alan Lessoff - Dept. of History	Records of the Journal of the Gilded Age and Progressive Era (JGAPE)	Society for Historians of the Gilded Age and Progressive Era	LOC Name	Both physical and digital materials were transferred to Archives at the same time. Digital materials were transferred from Alan Lessoff's home PC to a portable drive. Materials were then transferred to Archives SAN. Physical materials were placed in the aquisitions shelving area.	Collection contains operational records for the journal as well as author correspondence and associated files related to their article that would be published in the journal.	journal (approx the last 5 years) were not included in this acquisition; donor will send materials once he completes his transition away from the editor position with the journal.	Physical materials were generally good conditio well organized. Digital f loosely correspond to th physical organization.
Physical - North Storage Stack; Digital - MLB_Archives\Archives\DigArch\Johnson_Jan	Jan Johnson - Milner Library	Records of Milner Library; Janet Johnson Papers	Johnson, Janet	Local	Physical and digital materials were transferred approx the same time to Archives.	Most items are Milner related in scope and will likely be arranged into the existing Milner collection. Materials are predominately related to Milner public relations including exhibit info, events, and speakers.	appraisal decision maker with some guidance from the Univ Archivist. Most items were kept with the understanding that the Archives would do most of the weeding. Records still in use were sent to the appropriate departments. A copy of the digital files was sent to Michelle Kaiden for reference for future public relations at	Most of the physical mat are in good shape but w to be rehoused from the they were delivered in t sturdy storage boxes. Th files are in donor's origi will need heavy weedin

Image 1: Digital Archives SIP file

2. *Accession*

   Save copies in the MASTER and PROCESSING folders; create manifest folder and processing files.

Step 1: Scan files for viruses using a dummy (not connected to the internet) computer.

Step 2: Save two copies of the bulk files. One copy is saved to the MASTER folder and the other to the PROCESSING folder. The MASTER and PROCESSING folders will have sub folders with the collection's name – this is where the bulk files will be stored.

**NOTE* The MASTER folder is never to be touched and only used to make copies should the processing/access copies become unusable.*

Step 3: In the PROCESSING folder, create a MANIFEST folder. This folder will contain processing information generated during arrangement of the collection, collected notes about the collection, and checksum files.

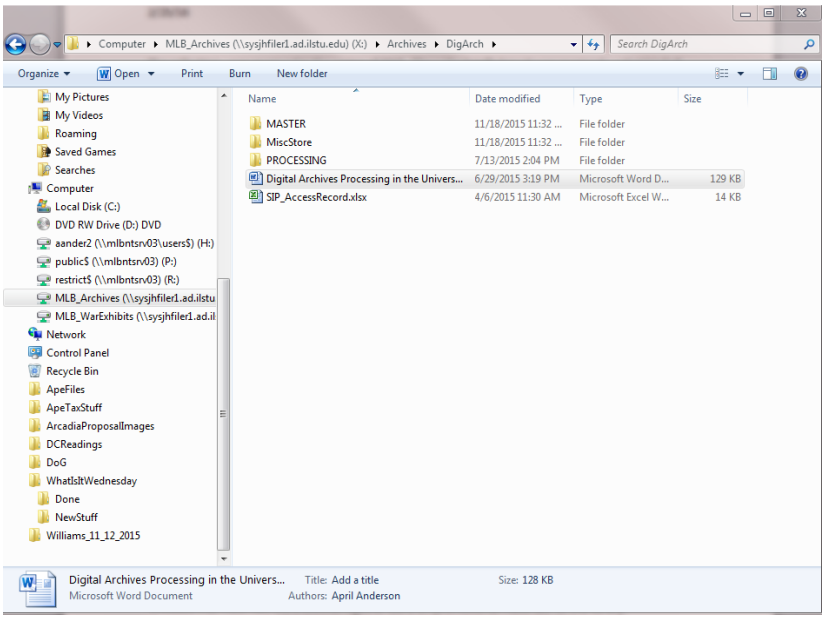

Image 2: Digital Archives MASTER and PROCESSING folders

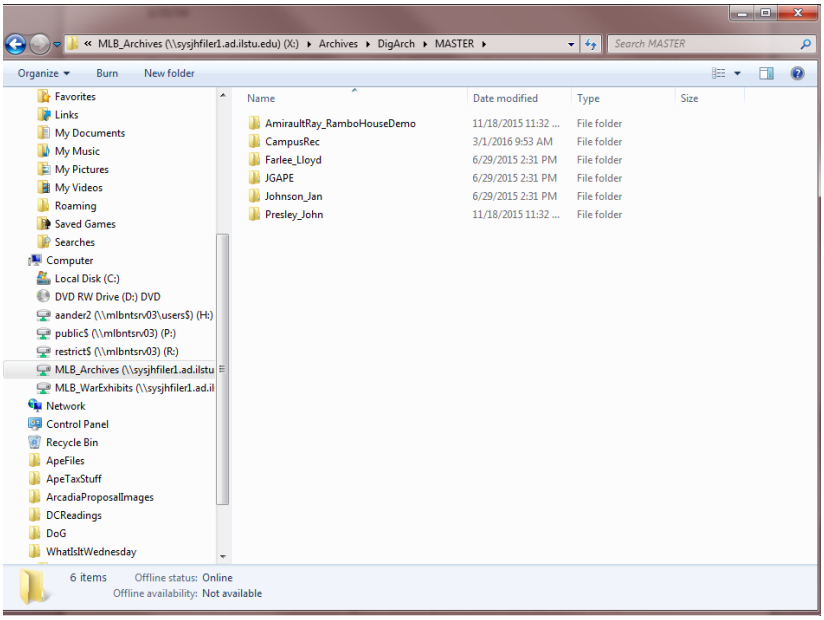

Image 3: The Digital Archives MASTER folder; these items are never opened and never altered.

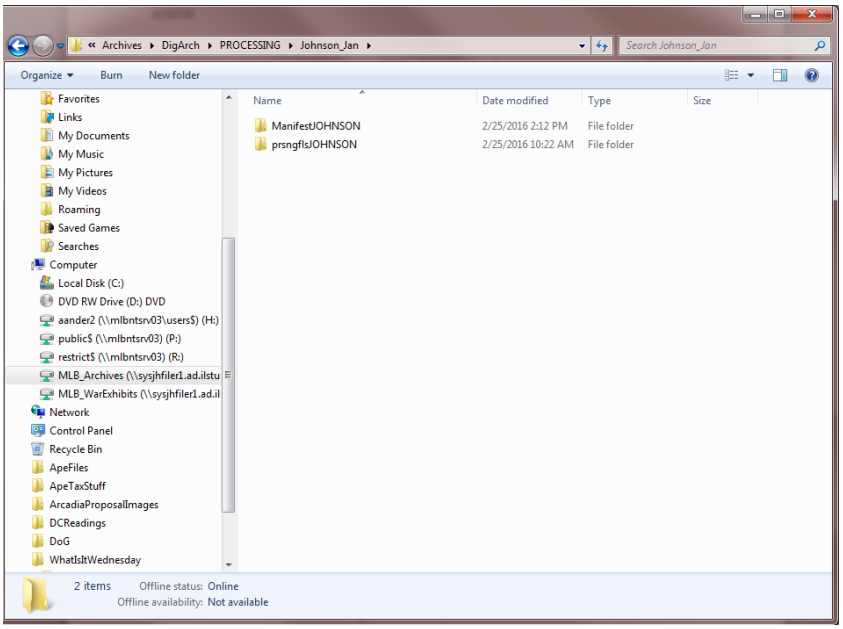

Image 4: The Digital Archives PROCESSING folder showing the archival collection of retired library staff member Jan Johnson.

3. Arrangement and Description

Step 1: Take an initial inventory using Karen's Directory Printer; create an overview file map and a detailed file list. Save these reports in the MANIFEST folder.

Step 2: Review the contents of the collection contained in the PROCESSING folder. Take notes on file types, large files, and any files that are corrupt and/or cannot be opened due to the lack of access to brand of software used to open certain files.

Step 3: Assess the software needs to view/edit files.

Step 4: After initial inventory and approval of hierarchy, begin sorting of materials into folders and removing copies and erroneous files. Give folders descriptive titles with underscored dates of the materials in the folders.

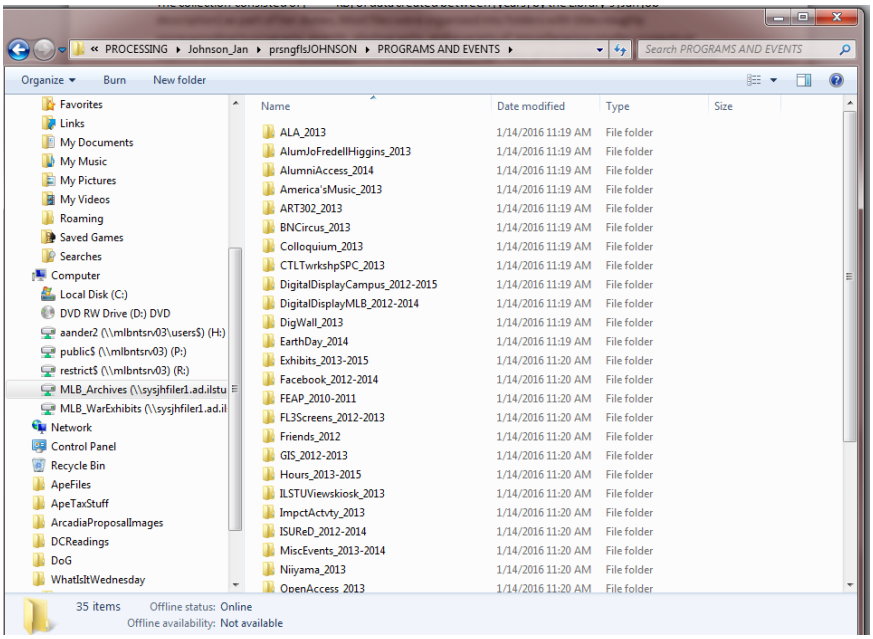

Image 5: The file structure of Series III in the Jan Johnson collection. This hierarchy will be translated into a searchable finding aid.

Step 5: As work progresses, remember to keep notes on processing work. Using NARA File Analyzer, take frequent checksums (once per week, on average) of the collection in process and place the reports in the MANIFEST folder.

Step 6: After the hierarchy is complete, make an ACCESS folder in the collection's PROCESSING folder. Place a copy of the processed collection in this folder. Normalize larger and non-standard files into accessible files, ie, TIFF files saved as lower resolution JPEG's and WordPerfect files as PDF-A. These will be used as access files by patrons.

Step 7: Write the discoverable finding aid. Include information on where to access the digital files.

Step 8: Perform a final checksum on the ACCESS folder. Move the contents of the ACCESS folder to the patron-accessible drive along with the checksum report. Use that checksum report as the baseline for future checksum reports on the ACCESS folder.

Step 9: Post finding aid.

4. *Discovery and Access*

Materials are ready for access by the patron. The access files can be made available to patrons by providing a link from the finding aid to the folder/file on a local server or cloud service. If using a local server, it is highly recommended to store the processed files on one server while the access files are stored on a different server.

Tiers Leading to Full Implementation

While it seems logical to invest now in a digital archives content management system, there are still many unknown factors in the university's digital archives needs. Though the Archives is collecting early data, they are still largely unaware of digital files and software systems being used by campus constituencies. Because of this, a campus-wide assessment is needed to determine the size, scope, and digital preservation needs of the campus community. However, the Archives cannot ignore the digital materials that are already being deposited. Those files include (but are not limited to) digital photographs, PDF's, Word documents, digital audio/visual formats, xml, html, and text files.

A tiered approach to implementation of a robust Digital Archives program at Illinois State University will not only allow the Archives to deal with immediate digital archives concerns but give us the information needed to select an appropriate content management software system that will address our as of yet unknown digital archives needs well into the future.

The following three tiers have been creating using information and recommendations from the white paper Digital POWRR.[2] More detailed information on the following tiers can be found in Appendix I and II.

Tier 1 – Open Source

Archives is currently testing within this tier to create a workflow, address campus and repository needs, as well as to model a digital archive that adheres to current national digital archives standards

[2] From Theory to Action: "Good Enough" Digital Preservation Solutions for Under-Resourced Cultural heritage Institutions. A Digital POWRR White Paper for the Institution of Museum and Library Services, August 2014.

and practices. The work already underway includes the utilization of proprietary, freeware, and open source software such as: Microsoft Office suite; IrfanView (to view graphics files but not to edit); Adobe (to view files) or complimentary software to create PDF-A; Karen's Directory Printer; NARA File Analyzer; university network drives (SAN servers).

The workflow currently being tested is piecemeal; various freeware and open source software are used on an as-needed basis. For example, before archives work can begin, a report of the drive's hierarchy and properties must be generated using specific software designed only for this task. As the archivist works through the collection, they take note of various files including documents, photographs, sound, and video. Each set of files will need specialty software to open and view the file's contents, ie files containing graphics must then be open with a compatible viewer. Some of the files in the collection are likely in a proprietary format and cannot be opened outright due to file structure changes and must first be opened as a PDF file. All of these tasks are performed separately and with different software that does not crosstalk with one another. While this process has certainly helped the Archives staff gain valuable insight into their work and the needs for a campus digital repository, this system is far from ideal.

As stated earlier, ISU's digital preservation capabilities are unclear. As such, developing any digital preservation workflows and standards during Archives tier 1 cannot be completed. Currently, all digital objects at Archives are being stored on a network of university computers known as a Storage Area Network (SAN), the parameters of which are unknown to Archives. In order to move forward on any tier level, a digital preservation repository must be acquired and tested.

However, Digital POWRR has a proposed tier 1 software option that Archives believes it can test now while a preservation repository is acquired. In their white paper, Digital POWRR propsed Archivematica as a compatible Tier I software package as it addresses our current needs and does so in a single software package. Archivematica offers new digital archives repositories a budget-friendly

solution for institutions looking to begin the process of developing a digital archives program. Use of the system will require open cooperation with the repository and their IT department. The system runs on a virtual machine that will likely need to be installed by IT and require some technical knowledge on the part of the repository's digital archivist. Should the institution choose, Archivematica offers per-hour contracts to assist in the installation, data migration, and training of staff. Advantages of Archivematica include:

- Open source
- Individual software for repository management versus multiple, independently maintained freeware and open source software packages
- OAIS compliant and compatible with multiple standards (including PREMIS and Dublin Core)
- Caters to both independent repositories and those who choose to pay for full service features including set-up, migration, remote hosting, and training.
- Allows for integration into other archives platforms including dSpace, CONTENTdm, and ArchivesSpace.

Archivematica is not a preservation repository; it only helps to streamline the multi-software process into one system that prepares the digital object for placement into a preservation repository. Though Archivematica does partner with DuraCloud, institutions can choose their own preservation repository.[3] With this flexibility and potential for added service enhancement, Digital POWRR has recommended Archivematica as both a Tier I and a Tier III solution.

Tier 2 – Hybrid

As Tier I does not test or report any digital preservation repositories, Tier II requires it. Regardless of the chosen archives processing workflow, there must be a digital repository in place to address preservation management and collection growth. Tier II suggest the use of a cloud based digital

[3] Brad Houston,"Archivematica," *The American Archivist Reviews*, last modified March 2, 2015, accessed May 10, 2016. http://www2.archivists.org/sites/all/files/ArchivesSpace.pdf

repository. By using a cloud service, file backup will be redundant and recoverable should a disaster on campus occur. The Archives will also be able to monitor file degradation and address problems directly as they arise through file monitoring and regular checksum schedules.

While it would be ideal to share management of a digital preservation repository with Administrative Technologies, obtaining clear information on current systems to help us manage existing collection has been difficult. With that in mind, this tier should also address and create a more fluid communication structure between Milner Library and Administrative Technologies. A contact person in the library should be selected and serve as the conduit through which questions and information about campus technology (and the technological needs of the library) flow.

Until a campus-based solution is found, Tier II suggests two services for cloud storage needs:

- Duracloud: already mentioned in Tier I as a digital preservation repository for Archivematica, Duracloud offers a one-copy depository with easy-to-use documentation. Users have online access to materials, automatic health checking services, and audio and video streaming directly from the cloud. The price point is high - $1,875 annually – but this service has proven to be reliable with other digital content management systems and would eliminate additional hardware like multiple streaming servers. Digital POWRR ranks Duracloud as both Tier I and Tier III software.
- Amazon Glacier: Amazon offers three tiers of storage: Standard, Standard – Infrequent, and Glacier. Duracloud utilizes Standard and while it would only cost the user $396 annually a year for 1 terabyte of storage directly through Amazon Standard, the user would only be allowed a predetermined number of upload/download requests and gigabytes of data transfer a month. The ‘slowest’ of the options, hence the term ‘glacier’, Amazon Glacier is a much cheaper alternative if immediate access to your stored digital files is not a concern. Amazon Glacier “is

optimized for infrequently accessed data where a retrieval time of several hours is suitable"[4] and charges approximately $132.00 per 1TB annually.

Finally, website archiving and ingestion of non-standard file formats such as GIS files will be investigated and researched within this tier. While the establishment of a digital preservation repository is imperative, the rapid evolution of campus websites and rise in the use of non-standard files need to be addressed once Archives has settled on a digital archives program.

Tier 3 – Vendor Service, Digital Archivist on Board

This tier will see the purchase of proprietary software and the hiring of a full time digital archivist. This tier will be developed as the result of the work done in tiers one and two as well as incorporate the information gathered on the university's digital archives needs. The chosen software will automatically handle many of the workflow steps developed through the tiers from ingestion to patron access to preservation.

As outlined above, Archivematica and Duracloud were considered Tier III compatible solutions by the Digital POWRR team. Open source and customizable, these systems each had their own price point according to the service needs of the customer. And while they streamlined the process already being tested at Archives, they each must be used in part with another system. Using a vendor based solution eliminates all need for multiple platforms and even provides the user with a digital preservation repository. However, pricing for vendor based digital content management and preservation solutions can be expensive.

Of the vendor based solutions, Digital POWRR recommended Preservica as a Tier III software option. Preservica automates all aspects of digital collection management from ingest to the repository, preserving the collection, and providing user access. The software allows for multiple platform ingest including most Microsoft products (including Outlook), Gmail, and even website harvesting. Preservica

[4] "Amazon Glacier", Service Information, last modified 2016, accessed May 10, 2016. https://aws.amazon.com/glacier/

has created a set of workflows that are OAIS compliant, should a user not want to develop a workflow from scratch. It also has unlimited support and offers daylong training of its product. Pricing for Preservica is as follows:

- Starter: up to 250GB on Amazon S3, $3950 annually
- Starter Plus: up to 500GB on Amazon S3, $6950 annually
- Standard: 1-10TB, first TB in Amazon S3 then $1,450 per additional TB per year and/or $550 per TB in Glacier per year
- Volume: 10TB+, no pricing listed online but website says this group has discounted plans

Though each system has its own qualities and characteristics, choosing a service and a workflow for a digital archives program at Archives will depend on the information gathered during the work being done in Tier I and II. Options presented in each tier can also be interchangeable among each tier. As the digital archives program is developed, it is important to always consider the safety and accessibility of the materials being collected. It is ultimately the Archives responsibility to provide preservation and access to the university's history. Using these tiers as guides and through inter-departmental and campus cooperation, the digital memory of the university will persist long into the future.

Conclusion

The creation of an infrastructure for the collection and preservation of digital files is critical to the success of the Dr. JoAnn Rayfield University Archives' mission. This infrastructure will rely on well documented ingestion and handling procedures, an analysis of ISU's digital preservation needs, and the appropriate hardware and software. By choosing the appropriate plan and addressing these issues as soon as possible ISU will minimize data loss, and continue to provide access to students, faculty, and scholars to our rich documentary heritage.

Appendix I

Glossary of Terms

1. Working with Files

    a. Access – the practice of facilitating access to physical and digital collections, both processed and unprocessed, in such a way that preserves their physical and intellectual integrity.

    b. Checksum/Fixity – a summary generated by a monitoring system informing the user of a the integrity of a given file or set of files. Checksums/fixity files are created from an algorithm (MD5 or SHA-256) which verifies that a copy made of an original digital record is identical to the original.

    c. Digital object – a file, system, or process that can refer to any type of information. A digital object can be a single text file or a file that requires child objects in order to function.

    d. Hierarchy – a detailed inventory of a collection, usually to a folder or item level.

    e. Manifest Folder – in digital archives, this folder is where information generated during archival processing of a digital collection is placed.

    f. Master Folder – in digital archives, this folder is a locked and monitored folder that holds original digital objects. The contents of this folder are never accessed and only used to create copies of files should the access copies become corrupt.

    g. Migration – the continuing process of transferring files of older formats to a newer equivalent.

    h. Normalization – moving files of a different format (usually proprietary) to a persistent format (PDF/A, XML, ASCII, etc).

i. Optical Media: media that contains bits recorded into a select material (typically aluminum) mounted to a thicker base (typically polycarbonate) and is both created and read by a laser. Formats include, but are not limited to CD's, DVD's, LaserDisks, MiniDisks, DIVX, Blu-ray, and others.

j. Processing – the practice of using standards and best practices accepted by the Society of American Archivists to acquire, arrange, describe, and provide access to collections of historic value.

k. Processing Folder – in digital archives, this folder is where copies of a digital collection's original files are placed. The contents of this folder are where processing work is performed.

2. Software and Programs

a. Adobe – also known as Adobe Systems Incorporated, a proprietary software company that developed numerous media creation and editing packages as well as popular document formats. In Digital Archives, programs most often reference are Photoshop, Dreamweaver, After Effects, Acrobat, and Acrobat Reader. File formats include portable document format (PDF), Shockwave (SWF), and Flash video (FLV).

b. IrfanView – freeware that allows the user to view and manipulate graphics on various Windows platforms. http://www.irfanview.com/

c. Karen's Directory Printer – freeware that quickly and efficiently gathers information of a given directory in a computer's drive and generates reports of that drive's folder hierarchy and properties. Users can determine what level of information they want in the report. The developer passed away in 2011 but links to the freeware can be found on the developer's Facebook page: www.facebook.com/KarensPowerTools

d. NARA File Analyser – similar to Karen's Directory Printer, this freeware allows the user to analyze contents of a given computer drive. However, this freeware also generates checksums.

e. PDF – a file format created by Adobe Systems that is intended to operate outside of any software, hardware, or operating systems. Originally a proprietary format, the format became an open standard in 2008.

f. PDF-A – a subset of the PDF format that embeds long term information and contextual properties of a file rather than linking to those properties, ex. fonts.

g. SAN servers – a network of connected computer storage locations, a Storage Area Network (SAN) can be a mix of hard disk devices and tape backup. Illinois State University uses a SAN network for a multitude of campus services (including storage of University Archives digital materials), however it is currently unclear the extent of the SAN's use on campus.

3. Digital Archives Standards

a. MADS - metadata authority description schema; developed to complement MODS and to use data in MARC21 format; used to describe and provide authority control for names of people, orgs, events, and terms; designed to be used independently but can work well with MODS

b. METS - metadata encoding and transmission standard - for digital objects; has five major sections: descriptive metadata, admin metadata, file groups, structural map, and behavior.

c. MODS - metadata object description standard - for information resources; "MODS is designed to allow the importing of existing catalog data that is in MARC21 format,

so it is especially valuable if there is legacy metadata in MARC format that needs to be moved to another metadata standard"

d. PREMIS - standard for preservation metadata (preservation metadata implementation strategy); usually used in conjunction with METS; defines relationships between digital preservation entities (but not intellectual entities) including objects, events, agents, and rights.

Appendix II
POWRR White Paper Vendor Recommendations and Pricing

Archivematica

- Open source
- OAIS compliant
- Compatible with METS, PREMIS, Dublin Core
- Requires installation of a virtual machine to host application locally (IT)
- Positive reviews if you can handle the virtual machine
- Freely available, open source, also fee-based services available (open pricing for these services)
- Tier 1 or 3
- **Archivematica pricing via https://www.artefactual.com/services/**

*These prices effective **January 1, 2015**.*

Service	**Description**	**Rates**
Software development	Requirements analysis and documentation, architecture and application design, coding, QA, AtoM theming, deployment, technical and user documentation. More info	$150 per hour fixed fee contracts / $130 per hour time & materials contracts
Data migration (AtoM)	Get your legacy descriptions into AtoM: data analysis and mapping, migration scripts, test import, QA, production import. More info	$150 per hour fixed fee contracts / $130 per hour time & materials contracts
Consulting	Capacity assessment, repository planning, digital preservation and access workflow planning, strategy documents, business cases, project scoping, pilot projects. More info	$150 per hour fixed fee contracts / $130 per hour time & materials contracts
Hosting	Full-service packages including remote hosting,	AtoM hosting / Archivematica

Service	Description	Rates
	data backup, technical support, software upgrades, software patches	hosting
Installation and technical support	Local Archivematica and/or AtoM installation, technical support, software patches and upgrades	AtoM tech support / Archivematica tech support
Training	Individualized online training for Archivematica and/or AtoM	AtoM training / Archivematica training

Duracloud

- Brokered cloud storage
- Easy to use with clear documentation
- Minimal metadata
- Open pricing
- Tier 3
- All Duracloud pricing comes with these standard features: Amazon S3 storage of primary copy of content; online access to all content; content sharing; web-based administrative dashboard; automatic content health checking services; storage reports and statistics; included bandwidth (up and down)
- **DuraCloud pricing via http://www.duracloud.org/pricing**

Subscription Plan	Features	Annual Price
DuraCloud Preservation The DuraCloud Preservation plan is ideal for institutions that wish to store one copy of their content in the cloud. *Subscription plan is available with storage between 1-5TB of content.* Example use case: Back-up preservation storage for a small amount of content	• Standard features	**(Storage in Amazon S3):** • $1,875 (subscription which includes 1TB storage) • $700 for additional TBs
DuraCloud Preservation Plus The DuraCloud Preservation Plus plan is best suited for organizations	• Standard features • Automatic synchronization of content between primary and	**(Storage in Amazon S3 + Amazon Glacier):** • $2,000 (subscription

Subscription Plan	Features	Annual Price
that wish to store two copies of their content in the cloud. *Subscription plan is available with storage between 1-5TB of content.* Example use case: Archival storage for a moderate amount of content	secondary storage providers • Choice of secondary cloud storage providers • Automatic file recovery between copies	which includes 1TB storage) • $825 for additional TBs **(Storage in Amazon S3 + SDSC):** • $2,875 (subscription which includes 1TB storage) • $1,400 for additional TBs
DuraCloud Enterprise The DuraCloud Enterprise plan is designed to meet the needs of institutions that wish to store one copy of their content in the cloud and need to provide a variety of individuals, departments, research groups, etc. access to a single DuraCloud account. *Subscription plan is available with customizable storage amount based on customer needs. Contact us for custom quote for storage beyond 10TB.* Example use case: Long-term access storage for a variety of institutional content	• Standard features • Media serving • Account management • Sub-account creation • Permissions and access controls • User management • Coming Soon: Shibboleth authentication -- available to Internet2 and InCommon members	**(Storage in Amazon S3):** • $5,750 (subscription which includes 1TB storage) • $500 for additional TBs
DuraCloud Enterprise Plus The DuraCloud Enterprise Plus plan is intended to assist organizations that wish to store two copies of their content in the cloud and need to provide a variety of individuals, departments, research groups, etc. access to a single DuraCloud account. *Subscription plan is*	• Standard features • Automatic synchronization of content between primary and secondary storage providers • Choice of secondary cloud storage providers • Automatic file recovery between	**(Storage in Amazon S3 + Amazon Glacier):** • $5,875 (subscription which includes 1TB storage) • $625 for additional TBs **(Storage in Amazon S3 + SDSC):**

Subscription Plan	Features	Annual Price
available with customizable storage amount based on customer needs. Contact us for custom quote for storage beyond 10TB. Example use case: Preservation and access storage for a large amount of institutional content	copies • Media serving • Account management • Sub-account creation • Permissions and access controls • User management • Coming Soon: Shibboleth authentication -- available to Internet2 and InCommon members	• $6,750 (subscription which includes 1TB storage) • $1,200 for additional TBs
Additional Storage For storage beyond 10TB, please contact us for a custom quote. The price per TB decreases for accounts storing more than 10TB.		**Contact Us**

Additional Systems

Curator's Workbench

- Runs on desktop before storing files in an institutional repository or dark archive
- Generates METS and MODS descriptive metadata elements
- Assumes a considerable amount of knowledge about metadata schemas
- No longer being actively developed, poor documentation
- Free, open source
- Tier 1

Internet Archive

- Very basic
- No access restrictions
- Stopped there, can't be used for our purposes
- Theoretically tier 2

MetaArchive

- Community-owned private LOCKKS network
- Provides dark archive storage

- Institution provides server space to host other institution's off-site materials, this is reciprocated
- Conforms to standards for a TRAC certified repository
- Picky about file names
- Open pricing
- Tier 2 with another system for archiving

Preservica

- Vendor based
- Users install one piece of software locally for uploading content into Preservica
- Clear documentation
- Can gather content via webcrawling
- Robust training (in person or remote)
- Good customer service
- Open pricing for cloud edition
- Tier 3