

# CSU Case Study

---

Final Draft February 2014

## Institutional Bio and Back Story

### Metrics -Institutional Level

#### Number of Students and faculty

Students-6,107*		Faculty and Staff	
	Undergraduate-4,618*		Full-Time-297*
	Graduate-1,489*		
			Operating Staff
	*based on Fall 2012		

### Endowment and budget

#### Endowment

~3 million

#### Budget

Operating Budget

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

### Carnegie Classification

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

### Colleges/Academic Programs Offered

- Degree-granting colleges
  - Arts and Sciences
  - Business, Division of Continuing Education and Non Traditional Programs Education, Health Sciences, Pharmacy
- Degrees offered
  - B.A., B.A. I.C.P., B.A. U.W.W., B.M., B.M.E., B.O.G., B.S., B.S.I.C.P., B.S.U.W.W., B.S. Ed., BSN, Ed.D., M.A., M.A.T., M.F.A., M.S., M.S.Ed, M.P.H., M.S.W, Pharm.D.

## Metrics-Library Level

### Size of Library

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

### Staff

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

### Budget

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

### Digital **preservation** technologies currently in use

- **Eloquent Archives**
  - Houses collection information including information regarding metadata and location information for collection. Additionally, this is the public catalog where we input/store records describing our archival collections. The catalog includes some digital files hosted therein. The system is hosted internally.
- **ContentDM**

- This is a content management system hosted externally by CARLI. The majority of our digital files (pictures) can be located here. The system is hosted externally.
- **ProQuest ETD**
  - A new process we have in place at CSU in order to provide access to our student's thesis and dissertations. Students have the ability to embargo access to their publication, but it is still stored on ProQuest servers (and a local copy is kept).
- **Internet Archives**
  - This system is linked to ContentDM. Some of our publications (yearbooks and historical books) can be found through here. A copy of the documents is also housed locally.

### Equipment

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

## Defining Moment

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

## Self Assessment Results

After taking a self assessment survey it is possible to see that the Chicago State University Library and the CSU Archives and Special Collections has done much to get digital initiatives created and providing access to these records. Collaborating with the CSU Library allows CSU Archive material to be accessible; however, our institution lacks resources to support digital preservation of these materials. For example, CSU Library staff limited time for digital preservation work due to commitments and previously assigned duties, our digital program would be able to flourish if we had a full-time staff committed to creating and managing the digital items we produce in the Archives. Additionally, more hands-on training for staff and faculty on the issues of metadata and digital preservation would make our institution well-rounded. The assessment also made it clear to us that the funding we receive for digitization and digital preservation comes from soft-money (grants) and we need to allocate more

appropriated dollars/budget toward this work, and resources could come in many forms including staff and funds to purchase the equipment and services.

The assessment also indicated to us that we need to be thinking more about our collections in terms of risk, specifically in relation to our audiovisual collections that are in our university collections and in the special collections. We have a project plan underway to convert the historically important video and audio material to a digital format in 2014-2015 so we will not lose this information. However, one of the issues that arose in the assessment is that we need to have a policy in which we specify what types of formats we will and will not maintain and which ones we are not; and how long we will maintain these resources.

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

## **Policy Gap Analysis**

### **Where we are**

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

### **Digitization**

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

### **Records Management**

Since the University Archives oversees the records management program, we are responsible for the electronic records for the campus. We have recently begun exploring the usage of Eloquent Records as an electronic management software option. In addition, the Archive department is working with the Legal department and Information Technology office to implement policies and procedures for university records

management as well as a method to capture and preserve the historically relevant information.

### **Workflow**

When we receive/acquire digital materials, we document the number of items (ex. 32 CDs, 1 floppy disk, etc.) and whether a label exists. However, we lack a systematic workflow or method of viewing each CD to do an inventory of its contents and to check for possible corruption. All of this is done on staff work computers, as we lack a standalone computer dedicated to these tasks.

### **Where we want to be**

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

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

### **The gap**

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

Our campus is supportive of collaborative models, so having a solution that includes this would be suitable for us. The library staff currently has various levels of understanding of what digital preservation entails; thus, it would be ideal if those who have had the opportunity to learn more about it would share their knowledge.

The recent Illinois SB 1900 open-access bill mandates that “each public university shall establish an Open Access to Research Task Force” by January 2014. Therefore, it is essential for each public university to commit to “preserving the archival version of the resource in its highest quality in accordance to collection development policy retention decisions. The bill has opened the conversation on our campus, and now that there is an impetus on this issue we can most likely receive financial support to address it and implement its practice.

## **Digital Preservation Policy/Program Proposal**

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

## **Mission and Scope**

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

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

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

## Digital Preservation Workflow

### Selection

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

### Acquisition

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

### *Content Types*

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

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

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

### Curation

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

## Archiving

The University Archivist, Special Formats Cataloger, Systems Librarian, and the Digital Initiatives and Web Services staff will be responsible for creating and managing Archival Information Packets (AIP) in preparation for storage, in consultation with individual collection managers to ensure appropriate metadata assignment and access levels, providing packet-level metadata, and uploading packets to the selected storage solution.

## Storage

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

## Retrieval

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

## De-selection

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

## Strategy

### Communication and Education

- **Administrators:** Senate Bill 1900 encourages resource allocation for open access and long term storage. Administrators need to be educated regarding the importance of said bill and its ramifications in order to make funding commitments. Education work can be done through one-on-one meetings, workshops, and presentations.
- **Colleagues who are stakeholders:** Office of Sponsored Projects, Faculty (need more outreach), Library subject specialists (more outreach), Graduate School (ETDs), and Campus Information Technology Department
- **Content providers:** University Archives, Record Management, Special Collections, Digital Theses and Dissertations, Library Faculty (Digital Projects),

Office of Scholarly Communications, University Units/Departments/Offices, and faculty (open access publications)

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

### Allocation of resources

- Human Resources: Not-yet existing but necessary positions in bold: **Scholarly Communications Librarian (interim), Outreach Librarian** (not yet hired). **Curator of Digital Collections. Assistant Archivist. Repository Developer(s).** University Archivist. Special Formats Cataloging Librarian. Systems Librarian, Web and Digital Initiatives Librarian.
- Budget: This will be determined based on salary for above positions and funding for a service and/or product. We will need budget for: On-campus redundant servers. Dedicated hard-funded allocation for digital preservation activities. Regular budget for local servers/storage space and redundant campus backup. Budget for subscription fees for geographically disparate offsite digital storage.

### Policy Review

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

### Content Types

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

### Lessons Learned

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

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

## Evaluation of Tools/Services Tested-Reflections and Recommendations

### Test Files

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

### Formats

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

### Recommendation Levels

Highly Recommend

Moderately Recommend

Do Not Recommend

### DuraCloud

#### Reflection

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

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

#### Recommendation- Moderately Recommend

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

## Curator's Workbench

### Reflection

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

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

### Recommendation- Do Not Recommend

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

## MetaArchive

### Reflection

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

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

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

### **Recommendation-Moderately Recommended**

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

## **Archivematica**

### **Reflection**

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

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

One of the system features is ICA-Atom, which the institution could use as their content management system as the DIP that is created can be deposited here. If one does not want to use ICA-Atom, then they could use their content management system. We use ContentDM, but we did not test the two together. We figured out that if files are not recognized by Archivematica, then they do not show up in ICA-Atom. One of the major problems we had was

that we could not get our imported metadata to be shown in the DIP or under the AIP when there was more than one file being processed.

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

### **Recommendation-Highly Recommended**

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

## **Preservica**

### **Reflection**

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

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

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

**Recommendation-Highly Recommended**

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

## Appendix 1: Standards and Best Practices

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

Relevant standards include:

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