

IWU Case Study

I. Institution Bio & Back Story

Illinois Wesleyan University <http://www.iwu.edu>

1312 Park St.

Bloomington, IL 61701

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

Metrics: Institutional Level

Number of students, faculty and staff [Facts 2013-14](#)

Undergraduate Students: 2009

Faculty: 183

Staff: 263

Endowment and budget:

Endowment: \$214 million (July 31, 2013)

Operating budget: \$90 million

Carnegie classification:

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

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

Degrees offered: BSN, BFA, BM, BME, BA, BS

Metrics: Library Level

Size of the library

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

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

Number of volumes: over 300,000

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

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

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

Staff

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

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

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

Budget

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

Digital preservation technologies currently in use

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

Defining Moment

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

II. Self-Assessment Results

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

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

We are hoping to implement a hosted version of ARCHON in 2014 to help us coordinate metadata for understanding and accessing our distributed collections. CARLI, the consortium that hosts the ARCHON instance we will be using, is discouraging members from viewing it as a repository for digital objects themselves.

Not everything that is digitized or born digital is accessible through online portals. Some materials are completely offline. Our collection development policies guide these distinctions. Objects in our CONTENTdm collections are completely open access; some parts of our IR are restricted to campus IP ranges but may be accessed from anywhere with valid credentials.

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

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

III. Policy Gap Analysis

A. "As Is" Model – where you are

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

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

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

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

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

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

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

C. Gap – what specifically is keeping your institution from achieving the "To Be" (money, staffing, technical infrastructure, buy-in, etc.)

Our real risk comes from the actions of others that I cannot control and feel I've made little headway in addressing. Our institutional history is in jeopardy due to default choices people don't even realize they are making. We need to develop a sense of urgency that this discussion is important enough to take time away from everything else that people are doing. Our institution relies too heavily on posting born-digital content to our website and does not consistently transmit the same copy to the archives or use file naming conventions that would make it possible to automate harvesting with web-archiving tools.

If we can get people on board to discuss these issues, the next sticking point will be to get individuals to change their content creation practices and take ownership of DP as part of their regular work. For units that generate content that will need full DP treatment, cost-sharing out of

diminished budgets will be the next hurdle. Until we reach that point, I don't believe we can discuss who should be curating university-wide content or what the specific workflows will be.

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

D. Digital Preservation Policy/Program Proposal

1. Mission/Scope

The mission of the program is to preserve and sustain long-term accessibility to the unique digital collections created, purchased or acquired by the library for which it has custodial or curatorial responsibilities.

In consideration of our institution's available resources, our preservation policy for digital content aims to achieve the following objectives:

- preserving and providing continued access to digital material, both born digital and digitized;
- ensuring that preserved digital materials are authentic;
- preventing damage and deterioration of the physical media by ensuring adequate environmental control;
- creating digitized copies in formats that make future migration possible; and
- changing the format of digital materials to preserve their intellectual content if necessary.

We will make efforts to take the needs and desires of our community (including digital content creators) into consideration when making preservation decisions; however, as long as the library retains sole responsibility for caring for this content, we assert decision making authority for

- formats we will accept as is and which ones we will normalize, if applicable, to protect the content;
- levels of preservation/protection needed (storage vs. long-term, bit-level management);
- access levels needed and who the gatekeepers are for off-line material; and
- preservation systems considerations (*e.g.*, hosted or not, end-user accessible or not).

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

i. Selection: Any member of IWU's community is encouraged to nominate digital content for preservation or analog material for digitization, but actual selection decisions will be based on collection development policies and in coordination with either a liaison librarian or the University Archivist. Digitization actions will be accomplished based on criteria that still need to be agreed on within the library.

ii. Acquisition: Once selection decisions are finalized, appropriate acquisition steps include collecting creator-provided metadata (detailed on an e-records transfer form).

- iii. Curation (*i.e.*, managing files and metadata): Without the availability of a defined preservation system, the archivist will continue to manage master files and metadata as before: in an Excel file. We have not consistently created checksums for locally created digital objects, and it seems this small step in the DP direction will be achievable.
- iv. Archiving: Minimally, collecting objects and metadata using Duke Data Accessioner can be used at this time. Collection into AIPs through the use of Archivematica will be investigated further.
- v. Storage: At a minimum, I'd like to establish a practice of creating one additional copy and storing it out of the library with all content the library has curatorial responsibility for. This will require ITS support and an added server.
- vi. Retrieval: Also system dependent, but it is not the intention of the library to promise that unmediated, online access will be possible for all content we curate. Our current content management systems are adequate for our community's needs today.

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

1. Communication & Education to the following (at a minimum) is the responsibility of the University Archivist at this point, but all library personnel need to be educated about our DP capabilities. Any person who has contact with our community may be asked to describe our services and should have a basic understanding of the difference between storage and preservation of digital content and things we do to increase access.
 - i. Administrators: The bulk of this work falls on Academic Affairs since both the library and IT are likely to be involved. If DP extends to the wider community, the Provost will need to be conversant on the issues so they can be discussed in the Cabinet. Even if Public Act [98-0295](#) (formerly IL SB1900) never touches private schools, it would be helpful for us to be aware of what competitors will be planning for.
 - ii. Colleagues who are stakeholders: Library and ITS staff are the primary group as of now. Library colleagues in collecting areas need to agree on minimal metadata types to collect and a means of transfer/storage. Workflows need to be established for these practices. ITS has numerous administrative demands but has been receptive to issues surrounding DP and made positive moves in recent years (*e.g.*, mirroring the campus website and expanding server backup practices off campus). We need more server space, though, specifically for the library's RAID.
 - iii. Content providers: Anyone on campus could be involved but the Communications Office (graphic designers, web content creators and all photographers) are of particular interest from an institutional history viewpoint. Faculty in our School of Music recently emerged as potential partners: their interest is in furthering access points for selected student, faculty and guest recitals, and they are receptive (in theory) to expanding their content capture practices to acceptable preservation formats.
 - iv. General marketing/PR: Audience includes BOT, Cabinet, Chairs, individual faculty, staff and students; for staff members who reach our extended community, Advancement, Alumni Relations and Admissions will also be important to strengthen ties with. ITS staff reach many users on our campus and have been advocates for good data management practices for years; continued partnerships in securing selected portions of the digital objects being created means educating ITS staff on the kind of content that rises to the archives' interests.

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

i. People – Not likely to expand beyond the archivist and one other person. Currently this is the Digital Projects Coordinator, but recent experiences with a Copy Cataloging staff member for ingesting archival material in the IR are encouraging. This work is added on to other duties and will require further negotiation and training if DP workflows are added. It seems likely that this person would replace tasks being handled, in part, by the Digital Projects Coordinator rather than augmenting them. Work study students should be trainable on some other workflows, too.

ii. Budget – Not likely to have a dedicated budget line unless a subscription service is deemed necessary. I am not recommending subscriptions at this time although DuraCloud seems most suited to our needs. DP selection decisions are my priority now.

IV. Lessons Learned

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

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

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

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