Shining a Light on the Digital Dark Ages

Meg Miner, Illinois Wesleyan University
Shining a Light on the Digital Dark Ages

Meg Miner
University Archivist and Assistant Professor

Faculty Colloquium
September 20, 2013
Why “Dark Ages”?

“Digital Dark Ages” is a term coined in 1997
• Need to care for humanities’ electronic heritage
• Recalls the efforts of monks in the Middle Ages
  – practice of copying texts and dispersing them saved knowledge that predated the Dark Ages

Connotes loss of knowledge or lack of intellectual growth due to a loss of understanding of what has come before.
Codex Sinaiticus
4th century mss. Bible, 20th century reprint
Why now?

• IWU’s Digital Dark Ages date from late-1990s
  – Our historical record is fragmented: servers, hard drives, external drives and institutional memory. Oh my!

• 2012 request for long-time administrator.

• 2011 request for 1920s student life.
Concerns for IWU Heritage

• Our Web site is our default archive, but
  – we are not consistent in how we post.
  – our ability to understand change is in doubt.
  – our file types are not stable for the long term.

• Which version of a record is the official IWU word on a subject?
  – If there is a discrepancy, how will it be resolved?
  – Where does the original live? Who cares for it?
    The archives does *if found* but some slip by.
Concerns for Heritage (cont’d)

• Content Management Systems
  – Unlimited ability to create and distribute
  – You are our monks...but not all on the same page

• Social Media & Third Party platforms
  – Issues about ownership, ongoing access, file types
Using the Web for promotion of our work is important, necessary and useful but we make assumptions about the safety of our records that are misleading.

We must start making decisions about what we value for our cultural heritage and secure that content in ways that will ensure its survival.
Shining a Light on the Digital Dark Ages

True Confessions of a Book- and Paper-Loving Librarian

Meg Miner
University Archivist and Assistant Professor

Faculty Colloquium
September 20, 2013
Assumptions in the Digital Age

Storage is cheap and everywhere!

Is the content secure/accessible? How? By whom?

If it is digital it will always be available!

Will it be useable? If software changes or files decay how will we know what is gone before it is too late?
Storage

• Storage options are cheap and everywhere!
Storage ≠ Preservation
Storage (cont’d)

• Storage options are cheap and everywhere!
  – But human intervention is not (and we are so busy that it is easy to put off)
    • Files need naming, tagging, organizing AKA metadata
  – Do your backed up versions all match?
    • Do you spend time synching devices? How much?
Don’t forget: all drives fail eventually

“Do you have a back-up plan?” by John K
http://www.flickr.com/photos/johnkay/5200871042/
Storage (cont’d)

Media needs to be replaced every five years (aka migration)

“Bit Rot” by The Joy of the Mundane

http://www.flickr.com/photos/mundane_joy/2316921309/
Storage (cont’d)

• Storage options are cheap and everywhere!
  – But human intervention is not (and we are so busy that it is easy to put off)
  – Do your backed up versions all match?
  – Remember: you get what you pay for with physical storage and in the Cloud!
Storage (cont’d)

Backups on cheap and/or fragile media

“Data deterioration” by Univ. of VA Library Preservation Services
http://www.flickr.com/photos/uvalibrarypreservation/8033085991/
Cloud providers *may not* = stable, consistent access

- What security measures do they promise?
- Do they back up anywhere? How will you know?
- Do they offer an exit strategy if they go out of business?
- If the person who set up the account dies, who has access?
Served ≠ preserved
Compressed files
Third party contracts
Homepages only on IWU servers

Where is Student Life?
<table>
<thead>
<tr>
<th>Blog Title</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Student Teaching</td>
<td>Just another Blogs at Illinois Wesleyan site</td>
</tr>
<tr>
<td>IWU SAAC</td>
<td>Student-Athlete Advisory Committee</td>
</tr>
<tr>
<td>ACTION RESEARCH CENTER at Illinois Wesleyan</td>
<td>Use the Navigation Menu to Learn More About ARC’s Connection to the Community!</td>
</tr>
<tr>
<td>The John Updike Society</td>
<td></td>
</tr>
<tr>
<td>IWU Anthropology Club</td>
<td>Information for students interested in anthropology</td>
</tr>
<tr>
<td>ITS @ IWU</td>
<td>Comments on Illinois Wesleyan University’s Information Technology: planning and operations</td>
</tr>
<tr>
<td>Hispanic Studies</td>
<td>¿Qué pasa? Otro blog de Illinois Wesleyan</td>
</tr>
<tr>
<td>NSLS at IWU</td>
<td>Welcome to the National Society of Leadership and Success</td>
</tr>
<tr>
<td>GREENetwork</td>
<td>Tips and updates from Illinois Wesleyan University’s GREENetwork</td>
</tr>
<tr>
<td>Around the House</td>
<td>IWU English Department News</td>
</tr>
</tbody>
</table>
Can’t crawl content

Student Life & Culture in the Cloud = History at Risk
Perceptions of Permanence

• “All that’s on the internet, it’ll always be there.”
  – Quality of what’s there? Who is responsible?
Perceptions of Permanence

Is it usable?

IWU Greek System page 1997

Perceptions of Permanence

Is it useful? Not without metadata!
Perceptions of Permanence

NOT always going to be there

Hrm.
Wayback Machine doesn't have that page archived. Want to search for all archived pages under http://www.iwu.edu/provost/ ?

IWU Provost
2007

http://web.archive.org/web/*/http://www.iwu.edu/provost/feb07/*
Perceptions of Permanence

“...cannot guarantee that they always will be....”

http://www.iwu.edu/disclaimer.html
Accessed September 14, 2013
Perceptions of Permanence

• “All that’s on the internet, it’ll always be there.”
• Digital files deteriorate.
They call it *Bit Rot*

“The stakes are high. Mistakes 30 years ago mean that much of the early digital age is already a closed book (or no book at all) to historians.”

*The Economist*
April 28, 2012
Files do deteriorate & with no sign until you try to open them

“And this is why you backup your files...” By npurmort
http://www.flickr.com/photos/npurmort/7158077037/
Perceptions of Permanence

• “All that’s on the internet, it’ll always be there.”
• Digital files deteriorate.
• Software changes make the ability to read file content uncertain over time.
Software and “Permanent”

Dear Enid:

Thankó verù mucè foò á lovelù luncheoi á imaginà yoõ no÷ ii Indiá anã wondeò ià y samã Indianó witè whoi wá talkeã yearó a werá quitâ remarkable— buô thâ political atomiã bombó foò Indiá witè eager— burni noõ tì kill anù insects® (Several haã silk so that no insect would get caught
Software and “Permanent” (1997)
Software and “Permanent” (2001)
Perceptions of Permanence

• “All that’s on the internet, it’ll always be there.”
• Digital files deteriorate.
• Software changes make the ability to read file content uncertain over time.
• Some online content should not be permanent.
  – “Official” files are superseded and old content not removed from the Web site.
Perceptions of Permanence

**Illinois Wesleyan: Alcohol Policy and Adjudication Guidelines**
IWU Philosophical Statement. PHILOSOPHY: A review of alcohol issues at IWU begins with guiding principals upon which policies and programs are built.
[www.iwu.edu/judicial/Alcohol_Policy.html](http://www.iwu.edu/judicial/Alcohol_Policy.html)

**Illinois Wesleyan: Student Handbook Policy and Regulations**
Campus Environment. Illinois Wesleyan encourages the entire community to engage in positive dialogue involving different life experiences. Intercultural ...

**Illinois Wesleyan: Judicial Affairs - Whistleblower Policy**
Whistleblower Policy. I. PURPOSE AND APPLICABILITY. The purpose of this policy is to set forth Illinois Wesleyan University’s policy on Board member, faculty, ...
[https://www.iwu.edu/judicial/policy/WhistleblowerPolicy.html](https://www.iwu.edu/judicial/policy/WhistleblowerPolicy.html)

**Illinois Wesleyan: ITS Usage Policy**
IT Usage Policy. Just as knowledge and discussion are at the heart of teaching and learning, so too are information and communications technology such as ...
[https://www.iwu.edu/its/about/itsusage.html](https://www.iwu.edu/its/about/itsusage.html)

**Illinois Wesleyan: Student Handbook Alcohol and Substance Policy**
Alcohol and Substance Policy. Illinois Wesleyan University endorses the Drug Free Schools and Campuses Act and the required provisions of that act may be ...
[https://www.iwu.edu/judicial/...Student-HandbookAlcoholPolicy.html](https://www.iwu.edu/judicial/...Student-HandbookAlcoholPolicy.html)

**Illinois Wesleyan: Sexual Harassment and Sexual Violence Policy**
Sexual Harassment and Sexual Violence Policy. Section 1: Statement of Purpose. Illinois Wesleyan University ("University" or "IWU") is committed to creating, ...
[www.iwu.edu/judicial/handbook/SexualHarassment.html](http://www.iwu.edu/judicial/handbook/SexualHarassment.html)

**Illinois Wesleyan: Judicial Affairs Federal Policies**
Federal Policies. Federal Drug-Free Schools and Communities Amendment Act. The federal government has enacted two phases of legislation aimed at ...
[www.iwu.edu/judicial/MandatedPolicies.html](http://www.iwu.edu/judicial/MandatedPolicies.html)

**Social Media Policy**
This policy provides guidelines for ...
[File Format: PDF/Adobe Acrobat](http://www.iwu.edu/writingguide/social-media-policy.pdf)
Two steps towards preservation

Making back up copies of digital objects vs.

Making back up copies of digital objects and storing them in different places

• The minimum recommendation

• Geographic distance = more protection
  – IWU Web site and networked servers are backed up to another city.
  – Specialized software like Digital Commons
Searchable, stable, migrate-able
Browse Journals and Peer-Reviewed Series

**Constructing the Past (History Department)**
A publication of the Nu Gamma chapter of Phi Alpha Theta and the Department of History at Illinois Wesleyan University
- Visit Journal

**Pseudonym Required (English)**
This website is home to the creative writing of student authors at IWU. At Pseudonym Required, our goal is to help students become better writers by providing feedback to the authors and offering them a chance for their pieces to be published. Our main focus is helping authors improve their writing, but our submission and revision process is also designed to give us an opportunity to showcase the creative talents of our fellow IWU students. All of the pieces published this year as well as our archives of past publications are located on this site. Take a look at some of our student writers' work!
- Visit Journal

**Res Publica - Journal of Undergraduate Research (Political Science Department)**
Res Publica - Journal of Undergraduate Research is a publication of Pi Sigma Alpha, the Political Science Honorary Society. It is funded by the Illinois Wesleyan University Student Senate.
- Visit Journal
Full digital preservation

Deciding which digital objects need to be saved, describing them, backing up in geographically distant places, \textit{and} monitoring the digital files for deterioration.

\textbf{4 Ds and 1 M}

Decision making + Description + Duplication + Distance + Monitoring = Digital Preservation
Official definition: LoC

“Digital preservation is the active management of digital content over time to ensure ongoing access.”

http://www.digitalpreservation.gov/about
“Combines policies, strategies and actions to ensure access to content that is born digital or converted to digital form regardless of the challenges of file corruption, media failure and technological change.

“The goal of digital preservation is the most accurate rendering possible of authenticated content over time.”

– Preservation and Reformatting Section, Association for Library Collections & Technical Services, a division of the American Library Association; http://www.ala.org/alcts/resources/preserv/2009def
Then & now: the BIG difference

DP is widespread, cultural heritage concern.
• Time elapsing between technological changes and scale of potential problems ahead.
  – Consider span of time between technological changes: handwriting to print to audio/visual media to digital media.
  – Consider the relative difficulty in creating earlier types of media and the proliferating paths to creating and storing that are available today.
• Attention to our needs now = survival.
A time travel survivor
Reasons to act
Hardware and media become obsolete

Chamber of Horrors [http://dpworkshop.org/dpm-eng/oldmedia/chamber.html](http://dpworkshop.org/dpm-eng/oldmedia/chamber.html)
Reasons to act, pt. 2

Loss of original file fidelity

Reasons to act, pt. 3
Intentional editing
Reasons to act, pt. 4

Quantity being saved and loss of context that gives digital objects their meaning

Channel the NSA...just a little!
“Active management” job opportunities

- **Archivists, Librarians & IT**: organize data, data mining, search/discovery processes
- **Computer scientists & Mathematicians**: digital forensics work on old media and obsolete systems, encrypting or breaking encryption
- **Lawyers**: intellectual property, contracts
- **Sociologists, anthropologists, psychologists**: hoarding as a phenomenon, coping with loss
- **Archivists**: decision making/selections, long-term planning for patron-base needs
In the beginning

• Analog media is okay alone for long periods
  – Changes are usually visible to the eye or discernible by the nose!

• My education in digital preservation (2007)
  – grant mandates, workshops
  – Bit Rot seemed of most concern; digital file degradation is not obvious until too late
Digital preservation solved

Figure F-1: Composite of Functional Entities
Digital Preservation at a Glance
The IMLS Grant

- No consortium (CARLI) or Big 10 help available.
- Do what librarians do best: collaborate!
- National Leadership Grant from federal agency awarded Fall 2011.

“Advises the President and Congress on plans, policies, and activities to sustain and increase public access to information and ideas.”
Objective

Investigate, evaluate and recommend scalable, sustainable digital preservation solutions for libraries with smaller amounts of data and/or fewer resources.
Projected outcomes

Reports! This is for a government agency, after all. (Final due May 2014)

• Case Studies
  – Interviews with faculty, staff, administrators.
  – Explore personal and institutional data management practices.
  – Compare results across institution types.
Interviewees

POWRR = 76 ; IWU = 15
Where do people keep their data?

- USB: 2
- Cloud: 2
- External Hard Drive: 9
- Work Hard Drive: 6
- University Network: 9
- Other: 1
- Tape Drive: 1
Where do people keep their data?

- **USB**: 2
- **Cloud**: 1
- **External Hard Drive**
  - **Staff**: 4
  - **Faculty**: 5
- **Work Hard Drive**: 6
- **University Network**: 3
- **Other**: 1
- **Tape Drive**: 1

**Legend**:
- **Green** = Staff
- **Light Green** = Faculty
Where do people keep their data?
Projected outcomes

Reports! This is for a government agency, after all. (Final due May 2014)

• Case Studies

• Analysis of digital preservation products and services
  – Ease of use, availability of support, costs
  – Two parts: Tool Grid and in-depth testing
## Projected outcomes

Evaluating tools and services available today

http://digitalpowrr.niu.edu/tool-grid/

<table>
<thead>
<tr>
<th>Digital POWRR Tool Evaluation Grid</th>
<th>Ingest</th>
<th>Processing</th>
<th>Access</th>
<th>Storage</th>
<th>Maintenance</th>
<th>Other</th>
<th>Cost</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACE (Audit Control Environment)</td>
<td></td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Free</td>
</tr>
<tr>
<td>AFF Open Source Computer Forensics Software</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
<td>Free</td>
</tr>
<tr>
<td>Amazon Cloud</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
<td>Varies</td>
</tr>
<tr>
<td>Archive-It</td>
<td>x</td>
<td>x x x x</td>
<td>x x x</td>
<td></td>
<td></td>
<td></td>
<td>Varies</td>
</tr>
<tr>
<td>Archivematica</td>
<td>x</td>
<td>x x x x x x</td>
<td>x x x</td>
<td></td>
<td></td>
<td>x</td>
<td>Free</td>
</tr>
<tr>
<td>BagIt Library</td>
<td></td>
<td>x x x x</td>
<td></td>
<td></td>
<td></td>
<td>x</td>
<td>Free</td>
</tr>
<tr>
<td>BagIt Transfer Utilities</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Free</td>
</tr>
<tr>
<td>BWF MetaEdit</td>
<td>x</td>
<td></td>
<td>x x</td>
<td></td>
<td></td>
<td></td>
<td>Free</td>
</tr>
<tr>
<td>Carbonite</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
<td>$149 – $599</td>
</tr>
<tr>
<td>Chronopolis</td>
<td>x</td>
<td>x x x x x x</td>
<td>x x x</td>
<td></td>
<td></td>
<td></td>
<td>$1,500 – $2,200</td>
</tr>
</tbody>
</table>
Projected outcomes

Testing five tools/services in-depth

Curator’s Workbench (UNC)

- File management costs depend on total file size (in TBs)
  - Remember: two copies of everything = minimum
  - Six copies of everything = optimum
Projected outcomes

Reports! This is for a government agency, after all. (Final due May 2014)

• Case Studies
• Analysis of digital preservation products and services
• National and local dissemination (throughout 2014)
  – Conferences and workshops
  – Web-based presentations
  – Meetings like this...thank you!
Unanticipated outcomes

POWRR and collaboration

• UK-based Project SPRUCE (Sustainable PReservation Using Community Engagement): creating a community-owned tool registry that will feed into the POWRR Grid.

• MetaArchive: Piloting technology to support small-scale models.

• NDSA (Library of Congress): Co-planning regional events for NDSA membership drives and POWRR workshops.
Observations

Local decisions are the most important first steps.

- We will not have an enduring historical record if we rely on http://www.iwu.edu or social media sites.
  - Lack of control/ownership, quality of file types and inconsistent metadata.
  - We must separate promotion from permanence. As of now, they must be two distinct things from a digital preservation perspective.
Observations

• Durability of digital objects is a factor of preserving the context with the file.
  – Metadata: significance of a digital object cannot be understood without it.

• We should not try to keep everything.
  – Some records do not need to remain digital.
  – Cultural/personal habits are against us here!
  – Records creators must be involved but your archivist can help!
Observations

Once we know what is important to keep and for how long, make decisions about

• what types of storage locations the digital objects need, who has access, who is responsible for maintaining.

• what type of long-term preservation system will work.
Decision making visualized

Good preservation decisions are based on a deep *understanding* of the *possible* content to be preserved.

Possible to preserve → Actually preserved → All Content
Recommendations for office and at home

• If you use specialized software, make sure to check old files in new versions that come out and save them as new.
  – Save a backup copy in a “standardized” format or one that you can migrate.
  – “Normalizing” to a known medium may still be the best hope for survival.
  – Your archivist is ready to help with decisions and format recommendations!
The best laid plans
Deciding which digital objects need to be saved, describing them, backing up in geographically distant places, and monitoring the digital files for deterioration.

4 Ds and 1 M
Decision making + Description + Duplication + Distance + Monitoring = Digital Preservation

For long-term preservation
Institutional Recommendations

What we can do in the short term:

• Create a record? Copy your archivist!
• Start a new group? Tell your archivist.
• Want to preserve student activities? Me, too! Let’s talk.
• Not sure how to start? Consult your archivist.
Institutional Recommendations

Harder questions

• What responsibilities should we take for content that is not on IWU servers?

• What about faculty data?
  – Consider investment in PAT cycle.
  – Consider external pressures like NSF & NEH grant rules for Data Management Plans.

• Solo system implementation not recommended.

Who will we partner with?

– How can responsibilities (staffing & systems) be shared—both on campus and with others?
Final Quiz

Digital Preservation is...

A. More than scanning and cloud storage.
B. Decision making that begins at the moment of creation.
C. A series of actions that will help secure personal and cultural heritage.
D. A shared responsibility for our institutional legacy.
E. All of the above.
Why should you care?

Thank You!

mminer@iwu.edu
x1538

Visit our project Web site
http://digitalpowrr.niu.edu

And see the Library of Congress’s suggestions about saving your own digital heritage...future archivists will thank you!

http://digitalpreservation.gov/personalarchiving