

OUR SHARED IMPERATIVE: PRESERVING DIGITAL CONTENT

Indiana Online Users Group
October 28, 2011

Carrie Daniels
University of Louisville Libraries

Taking digital preservation personally: horror stories

- Losing data – of a personal sort
 - ▣ At the hands of your computer
 - ▣ At the hands of a hacker
- Your digital “self” – planning for your online persona

What can we do now?

- Thinking about what you want to see survive you
- Think about file formats
- Think about metadata
- Backups, storage in multiple locations
- Investigating ways to share passwords securely
- Check out the Library of Congress' pages on preserving personal digital materials:
<http://www.digitalpreservation.gov/you/>

Why do I care?

- Archivists and historians rely on this kind of material to understand how life was lived and experienced in the past
- This kind of material is inherently fragile and must be actively preserved – benign neglect won't work

Digital preservation in our professional lives

- What's worth preserving?
 - ▣ Digital collections you spent time and money digitizing and describing
 - ▣ E-journals, e-books
 - ▣ Electronic theses and dissertations
 - ▣ Born-digital content – including webpages, tweets, and blogs
 - ▣ Scientific data

Digital preservation in our professional lives

- It's a scary world for digital files
 - ▣ Fires and floods
 - ▣ Bit loss
 - ▣ Hardware failure
 - ▣ Media obsolescence
 - ▣ Software obsolescence
- You do have some options...

Preservation Services/Solutions



Sorting these options out

- Different technologies:
 - ▶ Examples: LOCKSS, COPPUL, MetaArchive, PeDALS, and ADPN all share LOCKSS software
- Different philosophies:
 - ▣ Embedding infrastructure and knowledge locally (example: Hathitrust)
 - ▣ Operating as an external, third-party solution (example: OCLC's Digital Archive)
- Different methodologies:
 - ▣ Centralized box on-site (example: Tessella)
 - ▣ Centralized box off-site (examples: Portico, OCLC)
 - ▣ Distributed and off-site (examples: LOCKSS)

Sorting these options out

- Different approaches to access:
 - ▣ Open access – providing access to the content is part of the plan – example: Hathitrust
 - ▣ Mix of open and closed – provide access to some materials – example: LOCKSS
 - ▣ Access limited to content producer – examples: MetaArchive, Chronopolis
- Different approaches to file formats:
 - ▣ Format agnostic
 - ▣ Accept a limited list of formats
 - ▣ Some normalize and migrate on ingest
 - ▣ Some migrate when format is endangered
 - ▣ Some migrate on demand or when necessary

Sorting these options out

- Different approaches to versions
 - ▣ Some save the original plus any migrated files
 - ▣ Some save the migrated file only
 - ▣ Some save multiple versions of updated files
- Different relationships with content creators:
 - ▣ Some work with publishers to clear content for ingest by members/subscribers
 - ▣ Some work with member-supplied content

Choosing your option(s)

- What you choose may depend on many factors:
 - ▣ Your philosophy
 - ▣ Your budget
 - ▣ Your content
 - Do you hold the copyright?
 - Do you at least have permission to preserve?
 - ▣ Your technological comfort level
 - ▣ You may choose more than one option!

Digital preservation in my library

- Our considerations
 - ▣ We are not the flagship institution and our budget reflects that even in the best of times
 - ▣ We were creating digital content that needed to be preserved – photos, audio, and ETDs
 - ▣ We needed something better than our IT department offered, which was CDs
 - ▣ We have a willingness to learn and to work hard; we have a lot of heart
 - ▣ We were not excited at the idea of outsourcing something so essential to who we are

Our solution:

The MetaArchive Cooperative

- Based on LOCKSS – lots of copies keeps stuff safe
- Started with six partners plus Library of Congress
- One copy of your digital files at each of the partner institutions
- Geographically dispersed
- Integrity checked on an ongoing basis via voting
- Dark archive: retrieval only by “owner”
- Format agnostic

The MetaArchive Cooperative

- TRAC audited –
 - Trusted Repositories Audit and Certification
 - MetaArchive conforms to all 84 of the TRAC criteria
 - MetaArchive operates according to the standards of a trustworthy digital repository across each of TRAC's three major areas of activity and concern (organizational infrastructure, digital object management, technologies/technical infrastructure and security)

Our solution:

The MetaArchive Cooperative

□ Membership types

□ Sustaining

- Voting representative on Steering Committee
- Serve on Content, Preservation and Technical Committees
- \$5,500 plus \$1/GB year (plus LOCKSS Alliance membership)

□ Preservation

- One non-voting representative on Steering Committee
- Serve on Content, Preservation, and Technical Committees
- \$3,000 plus \$1/GB year (plus LOCKSS Alliance membership)

□ Collaborative

- A category for consortia that join as a group
- One non-voting representative on Steering Committee
- Serve on Content, Preservation, and Technical Committees
- Collaborative membership has its own fee structure

MetaArchive: What we liked

- Technologically, we liked the LOCKSS concept
- Philosophically, we liked the idea of being directly involved in the preservation of our content
 - ▣ We like being involved in running the organization
 - ▣ Not that we like going to more meetings, but the meetings are very well-run and the collaboration is real
- Financially, it worked for us
 - ▣ Also encouraged us to start preserving journals
- The TRAC audit is nice, even if it came after we joined

The MetaArchive Cooperative

Membership now at 19, including Indiana State University and international partners in Brazil, Spain, and England



What about you?

- Journals and ETDs?
 - ▣ What are your rights?
- Digitized books?
 - ▣ Same question!
- Government data? Datasets?
 - ▣ High volume...
- Webpages? Blogs?
 - ▣ Harvesting as well as preservation questions

Digital preservation: It's up to us

- We have to be the people that make the case to our administrations that the time is now
 - ▣ May get a boost from federal requirements for scientific data preservation plans
- We have to start somewhere
 - ▣ Don't let the perfect be the enemy of the good
 - ▣ Make sure you are adhering to best practices in creating digital content
- We have to help each other
 - ▣ Collaborate and/or share information and experiences

Links, etc.

- My email: carrie.daniels@louisville.edu
- MetaArchive: metaarchive.org
- LOC personal digital preservation site:
<http://www.digitalpreservation.gov/you/>
- UofL Libraries presentations:
<http://louisville.edu/library/presentations/>

Bibliography

- Hacked! In the *Atlantic*:

<http://www.theatlantic.com/magazine/archive/2011/11/hacked/8673/>

- Cyberspace when you're dead, in the *New York Times*:

http://www.nytimes.com/2011/01/09/magazine/09Immortality-t.html?_r=1