

Digitization & Preservation Symposium 2010-08-25

- Twitter #digipres
- Panelists: [Sarah McHugh](#) (Montana Memory Project), [Amy Rudersdorf](#) (State Library of NC) and [Taylor Surface](#) (OCLC)
- Rate of information growth is 50% greater than capacity.
- By 2011, 50% of digital information won't have a permanent home.

Digital Preservation: Basics, Practices, Choices for Libraries By Taylor Surface

- Digital content has to be stored as data in a computer, delivered through a computer, have a reading device, etc. to have it show for the end user.
- Data must adapt to new delivery methods over time.
- Defining the frontier (1994)
 - Report: Preserving Digital Information
 - Major findings:
 - Responsibility: creators, publishers, owners
 - Long-term requires deep infrastructure
 - Need organizations capable of: storing, migrating, accessing
 - Need certification for trust (depositories)
 - Need legal rights and capacity for rescue.
- OAIS: our vocabulary for structure, roles, terminology (ISO 14721)
 - Open Archival Information System is used for archival and preservation of digital objects
 - Use the term “ingest”
 - Trusted Digital Repositories (CRL implementing)
 - Enabling preservations actions:
 - Registries of data formats: Pronom, GDFR—now UDFR
 - Data format verification tools: JHOVE, DROID
 - Makes sure that data is correct.
 - Metadata elements: PREMIS v2.0, April 2008
- Digital Preservation Framework: what do you need?
 - The ALA's [definition of digital preservation](#)
 - Technological Infrastructure
 - Digital Assets
 - Organizational Infrastructure
 - Resources Framework
 - All of these rely on each other to make the digital preservation work.
 - **Key concept:** who are you serving?
- Costs (JISC Study, Estimates from UK National Data Centers):
 - Acquisition & Ingest: 42%
 - Archival Storage & Preservation: 23%
 - Access: 35%
- How to preserve unique, local collections:

- DuraCloud
- Ex Libris Rosetta
- Internet Archive
- MetaArchive Cooperative
- OCLC Digital Archive
- Tessella
- Open Source software
- Blue Ribbon Task Force Report on Sustainable Digital Preservation and Access: http://brtf.sdsc.edu/biblio/BRTF_Interim_Report.pdf
- How long will a TIFF file last on an external hard drive:
 - How to make sure the hard drive is still readable?
 - What computer/display will be able to read the information?
- Tiff vs. JPG
 - Tiff is preferred because it is uncompressed.
- Is batch processing risky?
 - Problems are traceability of where the file originated from.

Providing Access: Digitalization and Long Term Preservation Efforts in Montana **By Sarah McHugh**

- Achieving Understanding
 - Facilitated discussions among major stakeholders.
 - Universities
 - Colleges
 - State Historical Society
 - Historical Museums
 - State Government agencies
 - **Dark archives:** Places that didn't want to be part of online access, but want to store their information digitally for preservation.
 - Identifying digitalization efforts for access vs. digitalization efforts for long-term preservation/access.
 - Agreement that statewide efforts should include both.
 - Gold CDs
 - A LOCKSS system (what is this?)
 - Digital archives outside the physical place.
 - Local storage.
 - Discussion of options for both.
 - Long-term preservation should be collaborative and developmental.
 - Decision on long-term preservation tool.
- Building Documentation
 - Statewide Digitalization Registry
 - Montana Memory Planning Documents
 - IMLS Connecting2Collections Grant/Montana Survey
 - Try to avoid information being digitalized twice.

- Assist any interested institution in joining the Memory Project
 - Equipment, Formats, Collection Named, etc.
 - What is your time scale for getting finished with the digitalization project?
- [CONTENTdm](#) training.
 - CONTENTdm is a digital content management system geared toward providing web-based access to digital content.
- OCLC Digital Archive is a service usable for storing and preserving digital original files and digital masters. It supports CONTENTdm and any other digital repository with simple, straightforward workflows.
- Making it possible for all to play
 - Partnership between major stakeholders in Montana
 - Have traveling laptop and scanners for borrowing
 - Partnership between local library and local museum/archive
 - Managed as a statewide project open to all.
 - Governing committee that assists the project.
 - Shared technical and training support.
- Next Challenges
 - Educating efforts
 - Identifying and prioritizing valuable and at risk local collections.
 - Defining the scope of the [Montana Memory Project](#)
 - Do we put everything in there?
 - Newspapers? Large drain
 - Historical? Contemporary?
 - Expanding funding sources and partnership
 - Increasing training and support
- Information on [Digital Photo Guidelines](#)
- [Digitalizing Images](#)
- [Guidelines for the Montana Memory Project](#)
- [The Best Practices Exchange](#): Digital access to government information
- Thesaurus for metadata: Getty or LOC

A flickr of Hope: Harvesting Social Networking Sites with Archive-It

By [Amy Rudersdorf](#)

- **Born digital content:** created and lives only online.
- Since 1987 in NC, the State Library is responsible for all State publications
- Records: can be destroyed after a certain point.
- Publications: have to be kept “forever”
- PDFs and other documents, and videos, are now born digital and posted to the web.
 - This makes it hard for the repository to keep up since these organizations don’t send the information to the repository.
- In 2004, the federal government funded two Web Harvesting projects:
 - [Web Archives Workbench](#)
 - Capturing Electronic Publications

- In 2007, State budget includes a line item for digital preservation at the State Library
 - Archivist is from the Internet Archives
 - Digital Information Management Program (DIMP)
 - Funds participation in ongoing Archive-It service with sister division, North Carolina State Archives
- In 2008, Social digitalization and social networking “takes off”
 - Obama, Perdue
 - Interest in e-governance, Web 2.0, social networking sites (SNS)
 - Outreach opportunity; communication forum
 - Documenting social change, social reaction to important events, and a cultural phenomena
- In 2009, NC selects Internet Archive’s Archive-It (AI), collocates all harvest from 1996 on, begins considering social networking issues.
- They have no money, but beta testing a lot.
- **Web harvesting:** using web crawling software to automatically copy and collect internet content such as website pages. (National Library of New Zealand)
- Tools to use:
 - Internet Archives’ Archive-IT (AI)
 - Web Archiving Services (WAS)
 - ContextMiner (SNS specific)
 - OCLC Web Harvester
 - Host own harvester/Ignore
- AI was able to get information from: Facebook, flickr, Twitter, and Youtube
 - Not everything on Youtube is available
- How Harvesting 2.0 Works:
 - Trial & Error, Quality Review, Collaboration
 - Talk with stakeholders about how everything will be saved including comments on websites.
 - Documenting detailed changes for partners.
- AI is interested in testing LinkedIn, Slideshare, Vimeo, and MySpace
 - Are they relevant? Should they be kept?
- WAS partners looking at delicious
- Challenges Remain
 - “proxy mode” viewing
 - Scoping the harvests
 - Privacy settings/logins
 - Capturing new/other social networking sites
 - Maintaining list of active NC social networking site users
 - Facebook keeps changing.