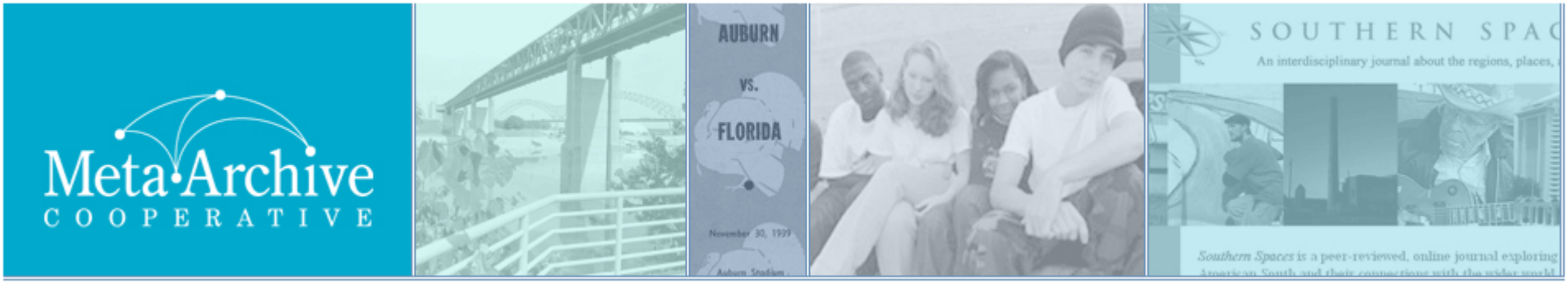


Distributed Digital Preservation: The MetaArchive Approach

Rachel Howard, University of Louisville Libraries

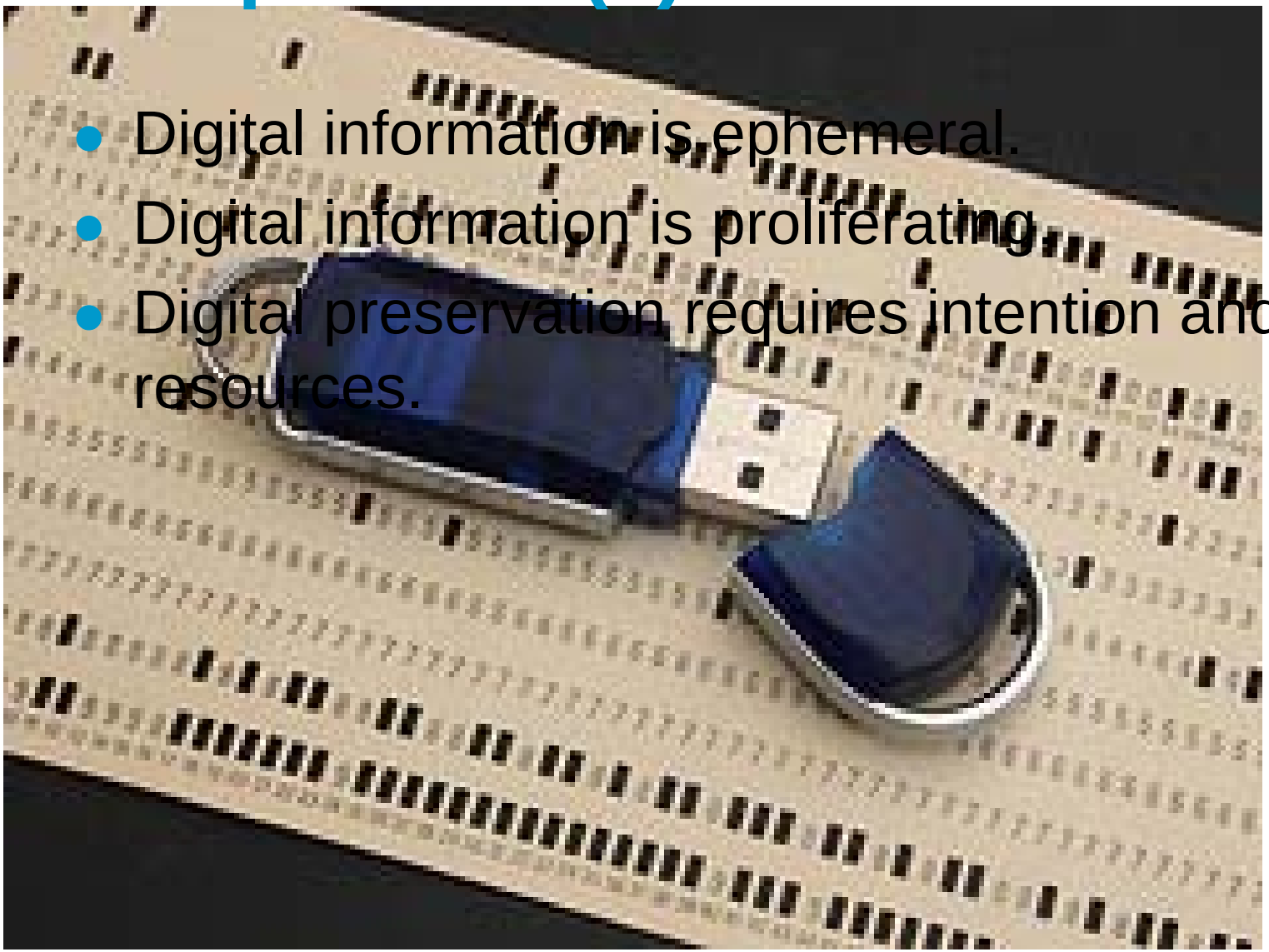
Ensuring Digital Access: A Forum on Digital Preservation

July 21, 2009

Overview

- Need for digital preservation/planning
- Concept of distributed digital preservation
- MetaArchive Cooperative
  - History
  - Private LOCKSS Network (PLN) elements
  - Organizational infrastructure
- How to join
- What's next

The problem(s)



- Digital information is ephemeral.
- Digital information is proliferating.
- Digital preservation requires intention and resources.

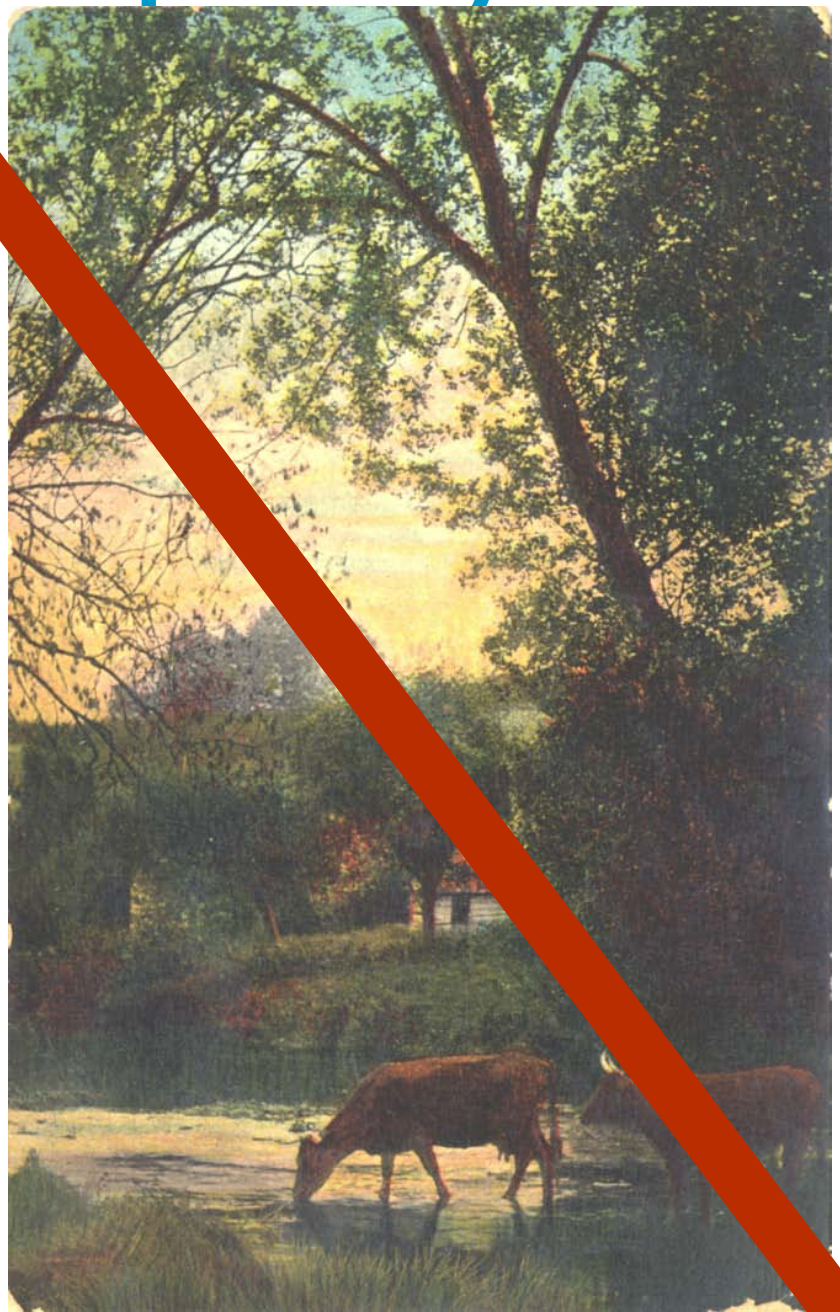
"Data storage – old and new." Made available under Creative Commons 2.0 Attribution License. Available at: http://flickr.com/photos/ian-s/2152798588/in/set-72157602236671297/.

THE GREATEST THREAT

to digital assets is not fire, flood or theft. It's the hazy assumption that cultural heritage institutions have taken the steps needed to preserve them.

Most often, we haven't. Which is why the MetaArchive Cooperative is leading a national effort to embrace distributed digital preservation, the future practice of digitally safeguarding the very items that define our culture and identity.

“Calf-path Syndrome”



Meta Archive COOPERATIVE

Best practices

- Save files in archival formats
  - Non-proprietary
  - Uncompressed (or at least not lossy)
  - In widespread use
  - Usable across platforms
- Make multiple copies
  - Preferably, have a copy on a server that is backed up.
  - Have another copy on Gold CD
    - Keep the CD somewhere distant from the server
  - External hard drives
- Keep technical and administrative metadata
- Implement a digital preservation plan
  - Collaborate for economies of scale

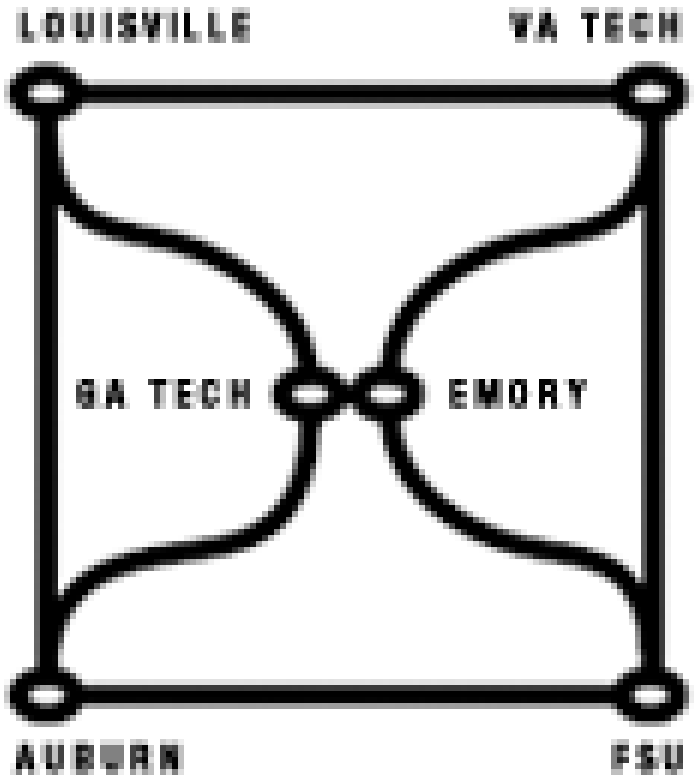
Components of collaboration

- Technical infrastructure
  - LOCKSS (Stanford)
- Organizational infrastructure
  - MetaScholar Initiative (Emory)
  - “Exposing Hidden Collections” conference (ARL)
- Financial infrastructure
  - National Digital Information Infrastructure and Preservation Program (NDIIPP)
  - National Historical Publications and Records Commission (NHPRC)

The MetaArchive of Southern Digital Culture

- Distributed digital preservation network for critical and at-risk content relating to the history and culture of the American South.
  - Develop a prioritized conspectus to ensure preservation of the digital materials most vulnerable to loss and in formats considered most at risk.
  - Use LOCKSS to harvest digital content from each other.
  - Create cooperative agreement for sustainable collaboration

Lots of Copies Keeps Stuff Safe



Private LOCKSS Network (PLN)

- Geographic distribution of preservation caches
  - Servers at each site harvest materials from every other cache, checking to make sure each copy is complete and valid.
  - Participants communicate permission to the LOCKSS system to harvest their materials via a Web crawler.
- Disaster recovery
  - A damaged cache can be re-built and re-populated from the identical sets of data in the other caches.
- Customized tool(s) to accommodate non-serialized content

Preparing items for harvest

- Define what is to be harvested
  - “Data wrangling”
  - Organize digital files into Archival Units (AUs)
- Provide access to the content and grant permission to harvest
  - Manifest pages (HTML)
- Tell LOCKSS what to harvest and where to find it
  - Plugins (Java)
- Notify partners to harvest new content

Archival Units

- One Volume of an Electronic Journal
- One Year of an ETD collection
- One Year of a scanned yearbook collection
- A folder of archival tif images or sound files

Manifest pages

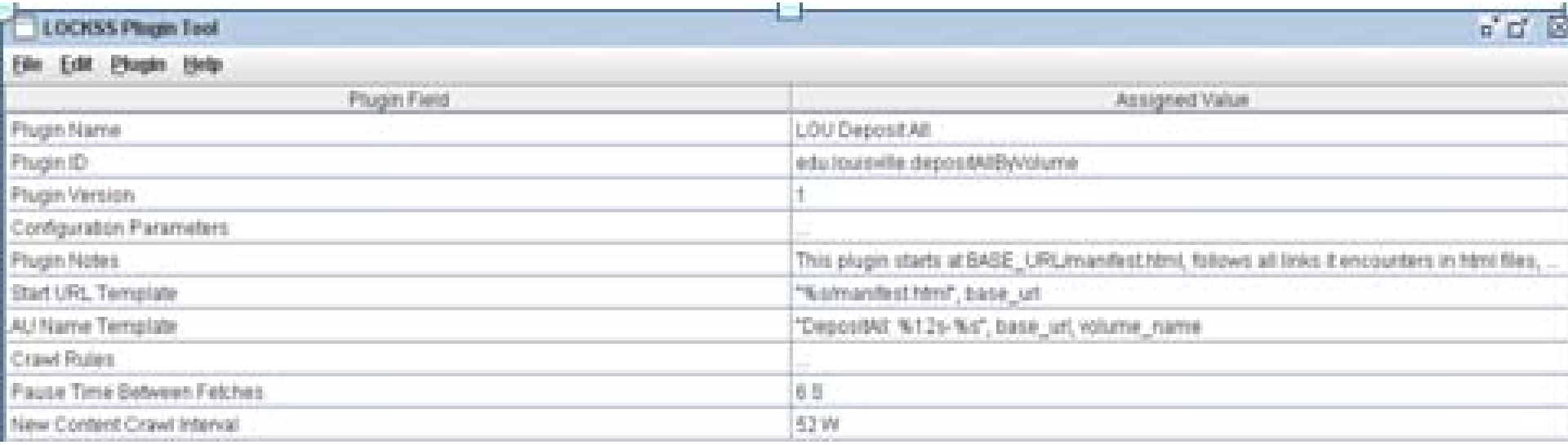
The Tin Horn LOCKSS Manifest Page

- 1925
- 1929
- 1930
- 1931

LOCKSS system has permission to collect, preserve, and serve this Archival Unit.

Plugins

- Instruct the LOCKSS software how to crawl and audit content.
- Plugin Repository accessed by LOCKSS Box upon startup.



Governance and sustainability

- Flexible organizational model
  - Charter
    - Broadly defines mission, goals, and activities of the Cooperative
  - Membership Agreement
    - Rights and responsibilities of members of Cooperative
  - MetaArchive Services Group
    - Nonprofit organization administers Cooperative
    - Minimal overhead

Committees and communication

- Steering Committee (meets in person 2/year)
  - Other committees have more focused charges:
    - Content
    - Preservation
    - Technical
    - Ad hoc
- Communication tools
  - Conference calls (1/week)
  - Listserv(s)
  - Wiki for document development
  - http://www.freeconference.com
  - http://doodle.com for scheduling meetings
  - http://oovoo.com/ for videoconferencing

Membership types and fees

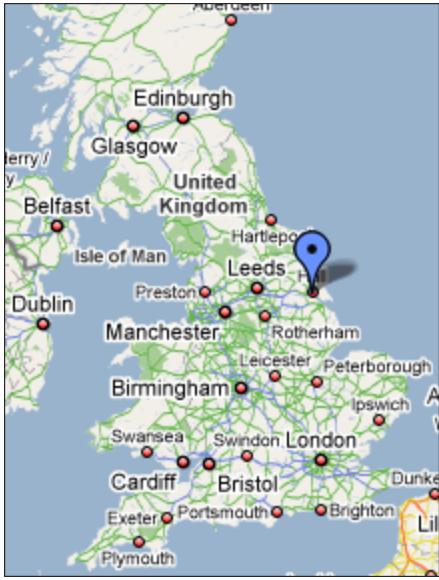
- All membership types presuppose membership in the LOCKSS Alliance (rates based on Carnegie classification) and a 3-year commitment
- Sustaining Members
  - Leadership role
  - Operate a node
  - Contribute more of content/year to be harvested
  - Cost: $5K/year or $12K/3 years
- Preservation Members
  - Operate a node
  - Contribute some content/year to be harvested
  - Cost: $1K/year
- Contributing Members
  - Contribute 1 collection/plugin to be harvested (can buy more)
  - Cost: $300/year

Expansion and adaptation

- Improving and expanding existing collaboration
  - Evolving standards and guidelines to offer as a model for new networks and collaborations
  - Enhancing technology, tools, and services
  - Wide applicability to a range of institutions and digital content
- Adding new networks
  - Electronic Theses and Dissertations Network
  - Transatlantic Slave Trade Network
  - Early Modern Literature

Membership distribution

12 US Members
+ Lib. of Congress



2 Overseas Members

Other collaborations

- Advising other Private LOCKSS Networks (PLNs)
  - Alabama Digital Preservation Network
  - Arizona Persistent Digital Archives and Library System (PeDALS)
- Ongoing exploration of projects to investigate and advance digital preservation
  - Data Preservation Alliance for the Social Sciences (DataPASS)
  - ECHO DEPository Project
  - SDSC Chronopolis

For more information…

- MetaArchive - http://www.metaarchive.org/
- LOCKSS – http://www.lockss.org
- NDIIPP - http://www.digitalpreservation.gov/