Webinar Roadmap...

Where we’re headed together this hour

Why we set off on this journey

Dodging the Memory Hole

Chronicles I (and II)

Future work
Johnson Signs Civil Rights Bill Into Law at Historic Ceremony

House Approves, 289-126

Laos Cease-Fire Plea Rejected as U.S. Trick
USNP to NDNP...

National Digital Newspaper Program

A partnership between the Library & the National Endowment for the Humanities

About the NDNP

The National Digital Newspaper Program (NDNP), a partnership between the National Endowment for the Humanities (NEH) and the Library of Congress (LC), is a long-term effort to develop an Internet-based, searchable database of U.S. newspapers with descriptive information and select digitization of historic pages. Supported by NEH, this rich digital resource will be developed and permanently maintained at the Library of Congress. An NEH award program will fund the contribution of content from eventually, all U.S. states and territories.

Learn more about the Library's role in the NDNP »

Program News

Check here for the latest news on NDNP and its website, Chronicling America.


» 08/29/2014: NEH Releases National Digital Newspaper Program Impact Study

» 08/02/2014: A Million Pages Now
But what if...?

- You’re not NDNP funded?
- You’re a smaller or less resourced institution?
- You want to start making incremental steps towards preservation?
Chronicles I: Objectives

- Chronicles in Preservation (2011-2014)
  - Building on the successes of the USNP & NDNP
  - Looking beyond high-quality digitized newspapers
  - What are the needs of legacy collections and born-digital newspapers?

http://www.educopia.org/research/chronicles
1. How can curators effectively and efficiently prepare their current digitized and born-digital newspaper collections for preservation?

2. How can curators ingest preservation-ready newspaper content into existing distributed digital preservation (DDP) solutions?

3. What are the strengths and challenges of three leading DDP solutions when used to preserve digital newspaper content?
Guidelines to Digital Newspaper Preservation Readiness
  • Documenting strategies from the essential to the optimal

Interoperability Tools
  • To effectively manage repository-to-repository exchanges

Comparative Analysis of DDP Providers
  • Describe DDP and document processes
Chronicles I: Partners

- Chronicles Committee (content holders & curators)
  - Boston College, Clemson University, Georgia Tech, Penn State, University of Kentucky, University of North Texas, University of Utah, Virginia Tech
- Advisory Group (professionals & practitioners)
  - Liz Bishoff (Bishoff Group), Robert Horton (IMLS), Sue Kellerman (Penn), Mary Molinaro (UKY), and Frederick Zarndt
- DDP Providers (repositories)
  - MetaArchive, Chronopolis, UNT-CODA
Chronicles I: Methodologies

- Partner Collection Surveys
- Partner Collection Samples
- Interviews
  - State Libraries (Wyoming State Library)
  - Newspaper Corporations (Dallas Morning News)
  - Journalists (Calgary Herald)
- Technology Development
  - Environmental Scan
  - Technology Inventory
  - Gap Analysis
- Data Exchange Tests
  - Preservation Readiness Plans

https://flic.kr/p/9z8dWi
Chronicles I: Collections

- **Types:** Digitized NDNP, Digitized Non-NDNP, Born-Digital Print, Born-Digital Web
- **Size Ranges:** 60MB to 7+TB
- **Formats:** TIFF, JPEG, JPEG2000, PDF, HTML
- **Metadata:** METS, MIX, MODS, Dublin Core
- **OCR:** XML, PDF, ABBYY, METS-ALTO, ALTO, PrimeOCR, etc.
- **Repository Types:** Fedora, DSpace, CONTENTdm, Olive ActivePaper, Veridian, custom repositories, simple webservers, hard drives, optical media, etc.
Chronicles I: DDP Sites
Chronicles I: Guidelines

- **Inventorying Digital Newspapers for Preservation**
  - How to record what content an organization has and how it is stored

- **Format Management for Digital Newspapers**
  - How to identify, validate, and migrate formats

- **Metadata Packaging for Digital Newspapers**
  - How to choose metadata formats, export metadata from repositories, and manage the storage of metadata

- **Checksum Management for Digital Newspapers**
  - How to generate and monitor fixity information

- **Organizing Digital Newspapers for Preservation**
  - How to structure folder hierarchies and names

- **Packaging Digital Newspapers for Preservation**
  - How to organize a collection for ingest into a digital preservation system

http://bit.ly/1qIDCGt
Comparative Analysis of Distributed Digital Preservation (DDP) Systems

Chronicling in Preservation Project

Authors: Matt Schulz and Katherine Skinner

2 April 2016

Chronicles I: Interoperability Tools

Index of /chronicles/GTWhistle

<table>
<thead>
<tr>
<th>Name</th>
<th>Last modified</th>
<th>Size</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Parent Directory</td>
<td></td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>bag-info.txt</td>
<td>27-Feb-2013</td>
<td>593</td>
<td></td>
</tr>
<tr>
<td>bagit.txt</td>
<td>27-Feb-2013</td>
<td>55</td>
<td></td>
</tr>
<tr>
<td>data/</td>
<td>27-Feb-2013</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>manifest-md5.txt</td>
<td>27-Feb-2013</td>
<td>148K</td>
<td></td>
</tr>
</tbody>
</table>

Bag-Software-Agent: bagit.py <http://github.com/edu/bagit>
Bagging-Date: 2013-02-27
Payload-Oxum: 529930395.2743
Contact-Email: jody.thompson@library.gatech.edu
Contact-Name: Jody Thompson
Contact-Phone: 404-894-9626
External-Identifier: GMASTech:1853/3790
Source-Organization: Georgia Tech Library
Organization-Address: 704 Cherry St, Atlanta, GA 30312-0900
External-Description: Faculty newspaper issues submitted by the Georgia Tech Library. Files were delivered as born digitally and scanned from print. License information and extracted text information is included.
Bag-Size: 526M

- **Bag Split/Unsplit Utility**
  - A tool to help with splitting and unsplitting large bags of digital content.

- **LOCKSSsum Validator**
  - This is a web service that has been produced for LOCKSS-based networks to compare a BagIt manifest with a list of corresponding checksums produced by a LOCKSS box for the purposes of ongoing audit.

- **Bad Filename Detection Tool**
  - This script can detect problematic system files

- **Bag Description Service**
  - This script automates the application of the FCLA Description Service to iterate over top of a set of BagIt-based content.

Since final publication in late March, the Guidelines have been downloaded as a PDF by over 400 users from 31 countries.

Since publication in early April, the Comparative Analysis has been requested over 300 times.

MetaArchive has actively incorporated the project Interoperability Tools into its preservation workflows.
...networks of people
System-wide change requires system-wide involvement
Dodging the Memory Hole

2014 forum
dodging the memory hole

jodna

November 10-11, 2014

http://www.rjionline.org/events/memoryhole
Key quotes, Missouri

- Journalism not only documents what’s happening, but also shapes those actions (Cliff)
- Backing-up is not archiving (Edward)
- The CMS is the librarian for a lot of newspapers (Cindy, @librynth)
- The archives of news media represent a community asset and should be made available as such (Mizell)
- We need to address how to keep an archive alive; it’s not enough just to keep it (Tom)
- You’ve made me smarter about ?s I need to ask... and scared the heck out of me, too (Teri)
Hazards:

- Definitions
- Copyright
- Revenue streams
- Technical complexity
- Ignorance
- Broken relationships
Common Refrains

- Big-picture Documentation
- Advocacy messaging
- Workflows
- Legal Deposit
- Multi-stakeholder Partnerships
- Additional Partners
Next Steps

1. Environmental scan
2. Born-digital PDF workflow/relationship documentation
3. Web archiving experiments (Archivelt, Storytracker):
4. Donation Agreement documentation
5. Outreach work, Advocacy
6. Public-private partnerships
Contact us!

Katherine Skinner
Executive Director, Educopia Institute
404-783-2534; katherine@educopia.org

Matt Schultz
Program Director, MetaArchive Cooperative
616-566-3204; matt.schultz@metaarchive.org