Collaborative Strategies for Digital Preservation of Newspapers

Martin Halbert
UNT Dean of Libraries
MetaArchive President

Monday, April 11, 2011
Newspaper Archive Summit
University of Missouri
Columbia, MO
UNT Texas Digital Newspaper Program (TDNP)

- 147 titles, 500K pages, 65K issues
- Will add 217 new titles and 100K pages in 2011
- 565K uses in 2010, 165K so far in 2011
- 5 FTE staff, 2 FTE students
- Funded by National Digital Newspaper Program (NDNP), LSTA, and (increasingly) local Texas foundations, newspapers, and private donors

[Logos for various funders and partners]
New TDNP focus on born-digital newspapers
Initial test ingests from UNT daily
Scaled up work with current newspapers from Rusk and Abilene
Building relationships through the Texas Daily Association and Texas Press Association
Beginning to better understand differences in ingest processes from digitization workflow
UNT Findings

- Digital newspaper program can successfully share same infrastructure developed for other UNT digital library efforts (CODA)
- Newspapers are of great interest to historians and other scholars
- Need to mainstream the digital newspaper staff (transition them from project-funded to library program-funded status)
- Need to better understand preservation issues and solutions for digitized and born-digital newspapers
The MetaArchive Cooperative:
Distributed Digital Preservation for Cultural Memory Organizations

- Established in 2003 under the auspices of and with funding from the National Digital Information and Infrastructure Preservation Program (NDIIPP) of the Library of Congress
- 384 TB distributed digital preservation network, organized as a nonprofit cooperative for libraries and other cultural memory agencies (research institutes, archives, etc.)
- Sustained by cooperative fee memberships, LC contracts, and other sponsored funding
- Provides training and models to foster broader awareness of distributed digital preservation and to enable other groups to establish similar networks
Other MetaArchive Digital Newspaper Collections

- Virginia Tech: campus, Virginia area newspapers, and some international newspapers
- University of South Carolina: NDNP content
- Penn State: campus and many regionals
- Rice University: campus newspapers and pamphlet series
- Georgia Tech: campus and others
MetaArchive Findings about Digital Newspaper Content Preservation

- Need for:
  - Additional guidance beyond NDNP for collaborative digital preservation of newspaper content
  - Techniques for ensuring file format viability
  - Strategies for organizing newspaper content consistently
  - Systems interoperability

- Most interest in:
  - Determine and share newspaper preservation readiness practices
  - Are there different issues to be addressed between preserving digitized and born-digital newspapers?
  - How can the Cooperative effectively track and manage newspaper collections?
  - Learning more about appropriate newspaper metadata.
Chronicles in Preservation Project

- Research and development project to study, document, and model the use of data preparation and distributed digital preservation frameworks to collaboratively preserve digitized and born-digital newspaper collections
- Participants: Educopia Institute/MetaArchive (lead), with the San Diego Supercomputer Center and the libraries of University of North Texas, Penn State, Virginia Tech, University of Utah, Georgia Tech, Boston College, and Clemson University
- NEH 2-year project ($300K award)
Guidelines for preparing digital newspaper collections for preservation

Interoperability Tools to facilitate the exchange of these newspaper collections between repositories,

Comparative Analysis of the strengths and challenges of three distinct DDP frameworks when they are used for the preservation of digital newspaper content
Conclusions

- Digital newspapers are a new and significant form of cultural memory
- Ben Franklin: “By failing to prepare, you are preparing to fail.”
- As a field we must prepare to work together to preserve our culturally significant content