Building a National Agenda for Saving Online News

Edward McCain, University of Missouri
Peter Broadwell, UCLA
Lisa Zirk, University of Missouri
Katherine Skinner, Educopia Institute
Roadmap

• Why online news needs saving
• What we’re doing about it
• How are we doing to date?
• Building a National Agenda
Why online news needs saving
Why online news needs saving
Why online news needs saving
Why online news needs saving
Why online news needs saving
Why online news needs saving
Why online news needs saving
Why online news needs saving
Why online news needs saving
Why online news needs saving
Why online news needs saving
Why online news needs saving
system-level change can be orchestrated through the deliberate work of stakeholders across the whole system.
What we’re doing about it

from act
to impact

@Educopia, Nov 6
What we’re doing about it

What we’re doing about it

from talking
to acting
(jump in!)
What we’re doing about it

2009 – LC/NDIIPP Workshop: Archiving Digital News
2010 – Citizen Journalists and Community News
2011 – The Newspaper Archive Summit
2013 – JDNA Launches
2014 – Guidelines for Digital News Preservation Readiness
2014 – Dodging the Memory Hole 2014 (Mizzou)
2015 – Dodging the Memory Hole: An Action Assembly (NEH)
2016 – Dodging the Memory Hole 2016 (IMLS)
Next...?
118 Attendees, 15 states+DC, two countries, LOTS of backgrounds
October 2016

2016 DtMH Goals:

- National Agenda
- Identify stakeholder roles
- Cooperation and knowledge sharing
- Examine news content management systems
- Prepare students
- Identify specific projects
What have we heard?

There is a crucial relationship between accountability and preservation (Hjalmar)

We need to ally tech/data infrastructures with archival techniques and principles (Kate)

Terms of service have changed. We are now renting and leasing, not owning in perpetuity. (Chris)

We have to be mindful of organizational stability—even strong corporations are not designed for long-term survival (Ana)

We need both business and social good arguments for preserving news (Matt)

We need to meet the nerds where they are. (Ben)

LC can demand copyright deposit. It has chosen not to. (Cliff)

We need to put archiving on the same footing as security, internationalization (Ben)

He who controls the past controls the future (Orwell via Evan)
What can we do?

- Create a registry of digital news preservation
- Collecting social media as primary source material
- Produce advocacy pieces designed for specific audiences
- Coordinate with associations serving each stakeholder community
- Create and test business models for preserving news
- Forward working models
- Produce more plugins for preservation for major platforms
- Assemble news hackers and lib/archive hackers (Hacks and Hackers meets Code4Lib)
- Talk to your local paper and report back.
Reflections from the event
Twitter collections at the UCLA Library

- Collected by Digital Library with Social Feed Manager
  – or –
- Collected by faculty with various tools and given to us later
- Examples: 785,000 tweets from Arab Spring movements (collected by faculty)
- Millions of tweets about various global disasters (Library)

The UCLA Broadcast NewsScape

- >306,000 hours of TV news archived digitally
- Recorded 2005-present, ca. 100 shows/day
- 14 countries, 11 languages
- 46 networks (20 non-US), 394,000 shows
- Searchable by ~3.75 billion words of captions, on-screen text, official transcripts.
Twitter/TV news linking experiment

in collaboration with Martin Klein, UCLA Library

• Apply DBpedia Spotlight Named Entity Recognition (NER) software to TV and Twitter collections on second GOP presidential primary debate on 9/16/2015
  • Twitter: 800,000 tweets
  • TV: CNN coverage of debate
  • Minute granularity
  • Persons, Organizations, Places

Results:
• Linked entities with URIs to DBpedia resources
• Visualization of correlations between entities
Twitter/TV News debate coverage: Persons

http://sologlo.library.ucla.edu/NER/twitter/gop_persons.html
Matching Twitter term profiles to news programs

17 Sep 2015 00:00 UTC: CNN CNN Republican Primary Debate, Anderson Cooper 360 matched 1526 terms

http://sologlo.library.ucla.edu/NER/tv/
Student Perspective; Student Involvement; Student Power!
How are we doing to date?