

# HERE TODAY, GONE WITHIN A MONTH: THE FLEETING LIFE OF DIGITAL NEWS

**Martin Halbert, Katherine Skinner, Marc Wilson, Frederick Zarndt**  
**Thursday, August 11, 2016**  
**Lexington, Kentucky**

**IFLA World Library and Information Congress**

82nd IFLA General Conference and Assembly



# NEWSPAPERS TRANSFORMED BY INTERACTIVE DIGITAL MEDIA

[www.loc.gov/ndnp/](http://www.loc.gov/ndnp/)

**FAVORITE HOMETOWN NEWSPAPER**

**- Since 1879**

## Researchers, publishers, and citizens ponder what this means



**Over the last two decades, U.S. newspaper publishers have transformed the news production cycle, moving from print to digital infrastructures and practices.**

Today, our most important news arguably is not delivered in analog form to doorsteps of subscribers daily, but rather through a wide variety of increasingly interactive media forms that include dense mixes of video, audio and text, as well as user responses.

For more than a century, libraries, archives, societies, museums, and other memory organizations dependably acquired, provided access to, and preserved print and microfilm news material for researchers.

In contrast, digital news sources today are stored on servers and accessed by readers via desktop and mobile devices, often with no physical manifestation or form.

# WHY IS DIGITAL NEWS PRESERVATION A PROBLEM?

[www.library.unt.edu](http://www.library.unt.edu)

**BEST HOMETOWN NEWSPAPER**

**- Since 1900**

## In the age of analog news, there was tangible content to preserve

**In the past, libraries and archives acquired news content through print (and later, microfilm) subscriptions, and then curated this hard-copy content without direct engagement with the news producers. This arrangement worked well.**

But unlike print newspapers, digital-only news has no physical form; instead, a subscription to digital content usually provides an institution or reader with rented, limited access to files that are managed by the newspaper producer.

There is a critical difference between this short-term **access** (viewing an object that is managed outside of the memory organization's own infrastructure) and **long-term acquisition** (actually bringing an object under the curatorial control of a memory organization).

**Under this newer access-based model, memory organizations most often do not take custody of the digital objects that comprise the “news” —** including image, website, social media, text, mobile “app”, and other content forms.



# BITTER IRONY: THE NEWS OF THE PAST WILL OUTLIVE THE PRESENT

[archive.org/web/](https://archive.org/web/)

**MARVELOUS HOMETOWN NEWSPAPER**

- Since 1920

## Ephemerality of digital news may lead to a dark age historically



Research has demonstrated that digital content of any genre/type that is not well organized and managed does not persist over time.

No accepted standard practice has surfaced for the way digital news content is produced or managed within this quickly evolving field. Rather, practice continues to be informed by decades of experimentation and steered by new content types and delivery mechanisms.

Surveys demonstrate that most U.S. newspapers maintain their digital news files primarily for near-term access (five years or less) not long term preservation.

**This leaves us with a bitter irony: today, one can find stories published prior to 1922 in the Library of Congress's Chronicling America and other digitized, out-of-copyright newspaper collections but cannot, and never will be able to, read many stories published online less than a month ago.**

# CASE STUDY: TOWNNEWS.COM

www.townnews.com

CONSUMMATE HOMETOWN NEWSPAPER

- Since 1960

## Struggling newspapers don't (or can't) pay for news preservation

TownNews.com BLOX software, CMS designed for newspapers, is **used by 59% of US daily newspapers**, including small (13K) to large (640K) circulation.

In order to minimize costs, publishers usually delete stories after 30 days (after that it costs them slightly additional to keep the story mounted in TownNews). In fact, **on average 50% of all pages published in 2015 by all publishers in TownNews were deleted after 30 days.**

**This amounts to more than 860K pages of digital newspaper content deleted in one year alone!**

Currently, TownNews is archiving unformatted versions of this deleted content at no charge for newspapers, but the primary mission of TownNews is not to archive and preserve digital news. If business conditions change or when the current CEO retires, TownNews's policies and practices with respect to archiving news may very well change.



# CASE STUDY: HUFFINGTON POST AND PROPUBLICA

[www.huffingtonpost.com](http://www.huffingtonpost.com)

**UNPARALLED HOMETOWN NEWSPAPER**

**- Since 1980**

## All digital newspapers are especially ephemeral and fleeting



**Homepages of digital-only news publishers change even more frequently than hybrid print/digital newspapers.**

A study of the Huffington Post and ProPublica in June 2016 demonstrated that the homepages of these services changed at least every 2 hours, and sometimes more frequently.

These rapid changes to homepages (much less “internal” pages) are not preserved in any known publicly accessible repository.

The Waybackmachine, the most familiar web archive, takes varying degrees of snapshots of these digital newspapers, with far more attention given to the Huffington Post homepage (almost every half-hour) than ProPublica (or any of the TownNews sites).

**There are no comprehensive and systematic efforts to preserve the enormous number of rapidly changing online newspapers. We are in an era of uncertainty concerning long-term preservation of the digital news.**

# WILL WEB ARCHIVING SAVE NEWS?

[www.loc.gov/webarchiving/](http://www.loc.gov/webarchiving/)

**WORLDWIDE HOMETOWN NEWSPAPER**

**- Since 2000**

## Web archiving shows promise, but serious challenges remain

Archiving web-based content—including the diverse and ever-changing platforms and file formats involved—is a growing area of activity for many libraries and archives.

However, while the web continues its exponential growth and technical advances accelerate, web archiving program development at research libraries remains underfunded and fractionally staffed.

Research libraries and other memory groups that are building local expertise and systems to provide long-term curation for digital content are cautious about undertaking complex born-digital acquisitions, including new content.

They need models that clearly demonstrate how to acquire, manage, and preserve this content. They also need concrete data regarding the costs and benefits associated with this work so that they can argue effectively for funding these efforts locally.

The screenshot shows two web pages. The top page is the Library of Congress Web Archiving page, which includes a navigation menu with 'ASK A LIBRARIAN', 'DIGITAL COLLECTIONS', and 'LIBRARY CATALOGS'. It features a search bar and a 'GO' button. The main content area is titled 'Web Archiving' and includes a 'WEB ARCHIVING' sidebar with links to 'Web Archiving Home', 'About the Program', 'Cataloging', 'Collections in Progress', 'FAQs', and 'Contact'. The main text welcomes visitors and explains the program's mission since 2000. The bottom page is the CyberCemetery website, which describes itself as an archive of government websites that have ceased operation. It includes a search bar and a section titled 'End of Term Web Archive' with a search bar and a table of project background, partners, and data.

Project Background	Project Partners	End of Term Web Archive
The End of Term Web Archive captures and saves U.S. Government websites at the end of presidential administrations. Beginning in 2008, the EOT has thus far preserved websites from administration changes in 2008 and 2012, and is currently preparing for the 2016 election season.	United States Central Command Feb. 16, 2009	U.S. Department of State Official Blog Feb. 16, 2011

End of Term Web Archive	End of Term Web Archive	End of Term Web Archive
Number of websites captured: 2008: 3,305 websites 2012: 3,247 websites Terabytes of data captured: 2008: 16 TB 2012: 21 TB	Number of websites captured: 2008: 3,305 websites 2012: 3,247 websites Terabytes of data captured: 2008: 16 TB 2012: 21 TB	Number of websites captured: 2008: 3,305 websites 2012: 3,247 websites Terabytes of data captured: 2008: 16 TB 2012: 21 TB

# DODGING THE MEMORY HOLE

[www.rjionline.org](http://www.rjionline.org)

**NATIONWIDE HOMETOWN NEWSPAPER**

**- Since 1879**

## Efforts to Collaboratively Build 21st Century News Collections



The Dodging the Memory Hole series of workshops have brought together representatives from a variety of stakeholder communities for moderated conversations and planning sessions about how to jointly develop the capacity to transform born-digital news preservation practices field-wide:

Newspaper Archive Summit – Apr 2011

DtMH I – Nov 2014

DtMH II – May 2015

DtMH III – Oct 2016

These events have established points of alignment for management and preservation efforts nationally between news publishers, press associations, technologists, public libraries, academic libraries, journalism schools, archives, corporations, and funding agencies.

**Any strategies for sustaining digital newspaper content must start with catalyzing conversations and connections between these stakeholder communities.**



# THE CRISIS OF EPHEMERALITY FOR DIGITAL NEWS CONTENT

educopia.org

FAVORITE VANISHED NEWSPAPER

- Since ???

## Enormous amounts of news content are now lost every month

Today we find ourselves in a curiously troubling situation in which newspapers simultaneously have many more digital options for producing innovative new forms of content but are far less likely to preserve such content.

This crisis of ephemerality for digital news content presents the field with both a challenge and an opportunity to cultivate new models more suited to our evolving circumstances.

Web archiving is still a nascent field, however, with uncertainty concerning the long term roles, rights, and responsibilities involved in web content curation.

Perhaps the most important next step in addressing this crisis is for the field of stakeholders to continue seeking to understand the scope of the issues and the range of potential solutions. At least we have begun to acknowledge the importance of this crisis.



You aren't preserving the news for me??