Here Today, Gone within a Month: The Fleeting Life of Digital News

By: Martin Halbert, Katherine Skinner, Marc Wilson, Frederick Zarndt

Abstract:

In 1989 on the shores of Montana's beautiful Flathead Lake, the owners of the weekly newspaper the Bigfork Eagle started TownNews.com to help community newspapers with developing technology. TownNews.com has since evolved into an integrated digital publishing and content management system used by more than 1600 newspaper, broadcast, magazine, and web-native publications in North America. TownNews.com is now headquartered on the banks of the mighty Mississippi river in Moline Illinois.

Not long ago Marc Wilson, CEO of TownNews.com, noticed that of the 220,000+ e-edition pages posted on behalf of its customers at the beginning of the month, 210,000 were deleted by month's end.

What? The front page story about a local business being sold to an international corporation that I read online September 1 will be gone by September 30? As well as the story about my daughter's 1st place finish in the district field and track meet?

A 2014 national survey by the Reynolds Journalism Institute (RJI) of 70 digital-only and 406 hybrid (digital and print) newspapers conclusively showed that newspaper publishers also do not maintain archives of the content they produce. RJI found a dismal 12% of the "hybrid" newspapers reported even backing up their digital news content and fully 20% of the "digital-only" newspapers reported that they are backing up none of their content. Educopia Institute's 2012 and 2015 surveys with newspapers and libraries concur, and further demonstrate that the longstanding partner to the newspaper—the library—likewise is neither collecting nor preserving this digital content.

This leaves us with a bitter irony, that 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 a story published online less than a month ago. In this paper we look at how much news is published online that is never published in print

or on more permanent media. We estimate how much online news is or will soon be forever lost because no one preserves it: not publishers, not libraries, not content management systems, and not the Internet Archive. We delve into some of the reasons why this content is not yet preserved, and we examine the persistent challenges of digital preservation and of digital curation of this content type. We then suggest a pathway forward, via some initial steps that journalists, producers, legislators, libraries, distributors, and readers may each take to begin to rectify this historical loss going forward.

Keywords: news | born digital news | preservation | newspapers | e-edition

Article:

***Note: Full text of article below

2016 Satellite meeting - *News, new roles & preservation advocacy: moving libraries into action*
10 – 11 August 2016
Lexington, Kentucky USA, USA

Here Today, Gone within a Month: The Fleeting Life of Digital News

Martin Halbert
University of North Texas, Denton, Texas 76203, USA.
martin.halbert@unt.edu

Katherine Skinner
Educopia Institute, Atlanta, Georgia 30309, USA.
katherine@educopia.org

Marc Wilson
TownNews.com, Moline, Illinois 61265, USA.
marcus@townsnews.com

Frederick Zarndt
IFLA Governing Board, Coronado, California 92118, USA.
frederick@frederickzarndt.com

Abstract

*In 1989 on the shores of Montana's beautiful Flathead Lake, the owners of the weekly newspaper the
Bigfork Eagle started TownNews.com to help community newspapers with developing
technology. TownNews.com has since evolved into an integrated digital publishing and content
management system used by more than 1600 newspaper, broadcast, magazine, and web-native
publications in North America. TownNews.com is now headquartered on the banks of the mighty
Mississippi river in Moline Illinois.*

*Not long ago Marc Wilson, CEO of TownNews.com, noticed that of the 220,000+ e-edition pages
posted on behalf of its customers at the beginning of the month, 210,000 were deleted by month's end.*

*What? The front page story about a local business being sold to an international corporation that I
read online September 1 will be gone by September 30? As well as the story about my daughter's 1st
place finish in the district field and track meet?*

A 2014 national survey by the Reynolds Journalism Institute (RJI) of 70 digital-only and 406 hybrid (digital and print) newspapers conclusively showed that newspaper publishers also do not maintain archives of the content they produce. RJI found a dismal 12% of the "hybrid" newspapers reported even backing up their digital news content and fully 20% of the "digital-only" newspapers reported that they are backing up none of their content. Educopia Institute's 2012 and 2015 surveys with newspapers and libraries concur, and further demonstrate that the longstanding partner to the newspaper—the library—likewise is neither collecting nor preserving this digital content.

This leaves us with a bitter irony, that 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 a story published online less than a month ago. In this paper we look at how much news is published online that is never published in print or on more permanent media. We estimate how much online news is or will soon be forever lost because no one preserves it: not publishers, not libraries, not content management systems, and not the Internet Archive. We delve into some of the reasons why this content is not yet preserved, and we examine the persistent challenges of digital preservation and of digital curation of this content type. We then suggest a pathway forward, via some initial steps that journalists, producers, legislators, libraries, distributors, and readers may each take to begin to rectify this historical loss going forward.

Keywords: news, born digital news, preservation, newspapers, e-edition

1. Overview

Over the last two decades, U.S. newspaper publishers have transformed the news production cycle, moving from print to digital infrastructures and practices. During this time, this industry has struggled to reinvent itself to meet the changing expectations and needs of its audience. The early Internet, social media, and mobile "app" experiments by newspaper publishers help to chronicle the evolution of journalism, and as such, they are of great importance to our understanding of the early twenty-first century's changing business and technical environment. Likewise, the local, regional, national, and international news recorded within these digital files is essential for our national understandings of this intensive period of political, social, scientific, technical, and economic change. 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. Studies have repeatedly shown that these files are not produced with preservation or future usage in mind, and that they are highly unlikely to be maintained by their creators.[1]

---

[1] See Educopia Institute et al's 2015 survey of NC news producers and RJI's 2015 survey of US-based news producers, Skinner, Katherine and Nick Graham. "Scanning the Environment: North Carolina Born-Digital News Preservation Practices 2015." https://docs.google.com/document/d/1d4geifFZyhtSkvzk76BpzEmQOgkcpX4kz6RORpJH3Jg/edit and McCain, Edward. "Plans to save born-digital news content examined." *Newspaper Research Journal* 36, no. 3 (2015): 337-347. http://nrj.sagepub.com/cgi/reprint/36/3/337. See also Anne Wooten, "Preserving Born-Digital

Today's news is fleeting in form and substance.[2] Researchers will not have access to this crucial set of information unless action is taken now to collect and stabilize it.

The significant challenges in providing long-term access to digital news sources are both technical and social in nature, as described in more detail below. Numerous social complications compound these technical issues, particularly regarding the current and historical roles and responsibilities taken by different groups in "preserving" news content.

Rectifying these issues is crucial if we, as libraries, are to continue fulfilling our missions to preserve the cultural record[3].

2. Why <u>Digital</u> News Preservation Is a Problem

In the recent (print) past, news publishers and journalists produced and "broke" the news, usually through print runs and paper deliveries to subscribers' doorsteps. Libraries and archives acquired this content through print (or later, microfilm) subscriptions, and then curated this hard-copy content without direct engagement with the news producers. All stakeholders were served well by this arrangement. Newspaper producers focused on creating the news; memory organizations focused on preserving this historical legacy; and researchers had access to a broad range of content—national and local, subject-based and community-oriented—with little reason to fear for this historical record's disappearance.

The transformation of the news production process from print to digital dramatically impacted this acquisition and preservation workflow and the distribution of labor it enabled. Unlike print newspapers, digital-only news has no physical form that a memory organization acquires through a subscription. 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. There is no authoritative local "object" (or set of objects) akin to a physical print newspaper that memory organizations that subscribe to news content can automatically curate and preserve on behalf of the research communities that they

---

News at Digital Preservation 2014." *The Signal* https://blogs.loc.gov/digitalpreservation/2014/07/preserving-born-digital-news-at-digital-preservation-2014/
(All last accessed 2016-06-18).

[2] See e.g., http://blog.gdeltproject.org/the-incredibly-short-lifespan-of-an-online-news-article/ ; See also the recent work of Matt Welsh (LA Times) and Martin Klein (UCLA) on Twitter-based links and longevity issues, 2015-16, unpublished.

[3] Carner, Dorothy, Edward McCain, and Frederick Zarndt. "An international survey of born digital legal deposit policies and practices." IFLA International News Media Conference, Stockholm Sweden, April 2015: http://www.kb.se/dokument/utbildning/IFLA-KB-2015/13%20-%202015%20An%20international%20survey%20of%20born%20digital%20legal%20deposit%20policies%20and%20practices.pdf (Last accessed 2016-07-28).

serve. Ironically, the digital content is more useable to readers and researchers because it is searchable, because it can be accessed via the Internet, and because the contents can be indexed by search engines. For researchers, digital content is potentially much more useful than content stored in newsprint or in microfilm.

Consistently, research has demonstrated that digital content of *any* genre/type that is not deliberately organized and managed does not persist over time.[4] In other words, long-term access requires that digital content be actively prepared and cared for over time. Simply backing up poorly organized, non-version controlled files of highly diverse content does not adequately enable future access to valid and accessible content. By contrast, digital preservation (defined as the "series of managed activities necessary to ensure continued access to digital materials for as long as necessary") is now accepted as requiring organizational planning, policies, and actions well beyond what most repositories currently provide.[5]

News files created by leading industry players and start-ups, conglomerates and family-owned news groups alike range across the spectrum of digital file-types (.html, .xml, .php, .pdf, etc), referencing and connecting to an ever-changing constellation of databases, social media, and mobile app platforms. No accepted standard practice has surfaced for the way such news content is produced or managed within this quickly evolving field. Rather, practice continues to be informed by a decade-and-a-half of experimentation and steered by new content types and delivery mechanisms.

Further complicating this already fraught technical environment, "digital preservation" is not a part of the news industry's standard operations. A 2012 national survey conducted by Educopia clearly demonstrated that most U.S. newspaper respondents maintain their digital news files primarily for near-term access (five years or less), without seeking to ensure the longevity of such records. A 2014 national survey by the Reynolds Journalism Institute (RJI) of 70 digital-only and 406 hybrid (digital and print) newspapers concurred, with a dismal 12% of the hybrid newspapers having reported backing up digital news content and 20% of the digital-only newspapers having reported that they back up none.[6] Educopia's 2013 and 2015 surveys with newspapers and libraries demonstrated that libraries generally do not partner with news groups to collect or preserve this digital content. As journalist Meredith Broussard wrote in 2015 in *The Atlantic*, "I've been studying news preservation for the past two years, and I can confidently say that most media companies use a preservation strategy that resembles Swiss cheese."[7]

---

[4] See e.g., Sam Meister and Gabrielle Michalek, "Disasters at any scale: The MetaArchive Cooperative's community-based approach to risk mitigation and disaster preparation in the 21st century," IFLA, 2016: http://library.ifla.org/1357/1/083-meister-en.pdf (Last accessed 2016-06-08).

[5] See e.g., "Preservation Handbook Online," http://handbook.dpconline.org/ (Last accessed 2016-06-08).

[6] Dorothy Carner, Edward McCain, and Frederick Zarndt. "Missing Links: The Digital News Discontinuity." IFLA, 2014. http://www.ifla.org/files/assets/newspapers/Geneva_2014/s6-carner-en.pdf (Last accessed 2016-06-08).

[7] Meredith Broussard, "The Irony of Writing Online about Digital Preservation," *The Atlantic,* 2015-11-20: http://www.theatlantic.com/technology/archive/2015/11/the-irony-of-writing-about-digital-preservation/416184/

News files constitute a high-risk content genre, present significant preservation challenges, and are missing national-level preservation attention. 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.

3. Case Study: TownNews.com

In 1989 on the shores of Montana's beautiful Flathead Lake, the owners of the local weekly newspaper the *Bigfork Eagle* started TownNews.com to help community newspapers with developing technology. TownNews.com has since evolved into an integrated digital publishing and content management system used by more than 1600 newspaper, broadcast, magazine, and web-native publications in North America. Approximately 300 TownNews.com clients publish an e-edition[8] using TownNews.com services.
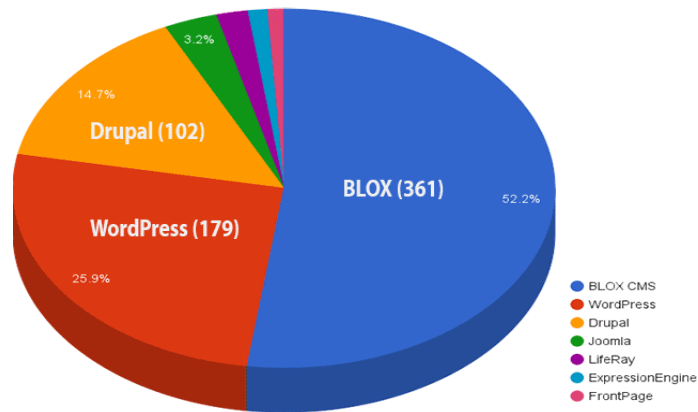
Why do we give particular attention to TownNews.com? According to a report[9] by Barrett Golding of the Reynolds Journalism Institute TownNews.com BLOX software, content management software designed for newspapers, is used by 59% of US daily newspapers. Because it is so widely used by news publishers, its policies and practices influence a wide range of US publishers, from small weekly news publishers like the Coronado Eagle & Journal (circulation ~13,000) to large circulation major daily news publishers like the St. Louis Post-Dispatch (circulation ~640,00).
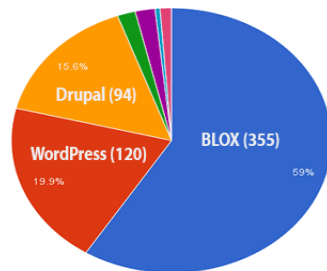
---

[8] e-editions are digital replicas of a newspapers' print edition.

[9] Golding, Barrett. The tools we use 1: Publishing print newspapers online: CMSs. Reynolds Journalism Institute. September 2015. https://www.rjionline.org/stories/tools-we-use-1-publishing-print-newspapers-online-cmss (Last accessed 2016-07-29).
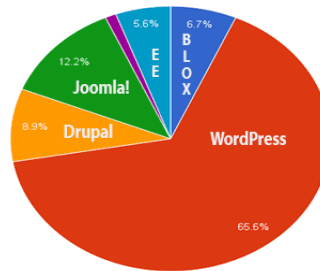
CMS share: U.S. newspaper sites (692)

Legend:
- BLOX CMS
- WordPress
- Drupal
- Joomla
- LifeRay
- ExpressionEngine
- FrontPage

BLOX (361) 52.2%
Drupal (102) 14.7%
WordPress (179) 25.9%
3.2%

Daily papers (602)

BLOX (355) 59%
WordPress (120) 19.9%
Drupal (94) 15.6%

Alternative papers (90)

WordPress 65.6%
Joomla! 12.2%
Drupal 8.9%
EE 5.6%
BLOX 6.7%

TownNews.com not only provides newspapers with online publishing support services; it maintains an archive --at its own expense! -- of past stories for born digital and digital only stories as well as for newspaper e-editions.

Why is this important? TownNews.com contract with e-edition publishers stipulates that a story is online for 30 days. After 30 days TownNews.com is no longer contractually obligated to keep it online, and, in order to minimize cost, publishers usually delete stories. The table below shows the number of e-edition newspaper pages published month-by-month in 2015 for TownNews.com clients and the number of these pages that are deleted after 30 days.

2015	Pages published	Pages after 30 days	% deleted
Jan	132199	78897	60%
Feb	123366	75914	62%
Mar	141621	84073	59%
Apr	140127	82284	59%
May	145633	78758	54%
Jun	141435	72758	51%
Jul	145294	67641	47%

Aug	149233	64565	43%
Sept	155047	66342	43%
Oct	153866	65187	42%
Nov	143305	59880	42%
Dec	146651	58747	40%

Newspaper publishers typically have no or very limited archiving or preservation policies and processes as was shown by Carner et al in *Missing links: The digital news preservation discontinuity*[10]. With no preservation or archiving policies and processes, the deleted pages would be permanently lost.

Fortunately for future generations CEO Marc Wilson thinks that archiving the news published by TownNews.com is "the right thing to do", especially in view of its clients' lack of archiving or preservation polices[11]. But TownNews.com is a business, one whose primary goal is to serve its customers and earn a profit. It's primary mission is not to archive and preserve digital news; if business conditions change or when Marc retires, TownNews.com's policies and practices with respect to archiving news may very well change. Archiving and preserving news is or should be the principal mission of other organizations, usually non-profit organizations or organizations in the public sector.

Other studies have similarly demonstrated the instability and fragility of digital news forms and infrastructures. As GDELT reported in 2015,

> Around 1.5-2% of the (news) articles we monitor on a typical day will no longer exist two weeks later, rising to 5% after a month and around 7-14% after two months. Just over 7% result in "404 Not Found errors", while another 7% yield everything from "500 Internal Server Error" responses to servers being offline or non-responsive (both of which could represent transient errors) to domains no longer existing (more frequent in certain areas of the world). Given that GDELT finds around 650,000 to one million news articles a day that it deems relevant, this means that 70,000 to 140,000 news articles a day are being lost.[12]

---

[10]Carner, Dorothy, Edward McCain, and Frederick Zarndt. Missing links: The digital news preservation discontinuity. http://www.ifla.org/files/assets/newspapers/Geneva_2014/s6-carner-en.pdf (Last accessed 2016-07-27).

[11] Without citing specific numbers or examples Marc Wilson reports that TownNews.com's clients do not have archiving or preservation policies or processes. Most TownNews.com's clients no longer have news librarians.

[12] http://blog.gdeltproject.org/the-incredibly-short-lifespan-of-an-online-news-article/ (Last accessed 2016-06-08). See also Scott Klein and https://www.propublica.org/nerds/item/a-conceptual-model-for-interactive-databases-in-news for an important examination of news databases and the preservation intricacies they introduce.

4. Case Study: Huffington Post and Propublica

The TownNews.com case study demonstrates that news stories often have a short life on the Web, at least on the Web that is indexed by search engines and visible to readers[13].

Another question one might ask is how often born digital news stories change during their life regardless of their lifespan. To show how often digital news stories can change, the authors tracked the homepages of digital only news publishers The Huffington Post (US edition) and Propublica using the web service Follow That Page!. We tracked both homepages for over a year, but here we show only the changes detected for the week June 23-30 2016.

Follow That Page! is a subscription web service that detects changes to web pages. The authors used the free version of Follow That Page! (there is also has a paid version). The principal difference is that the free version monitors a webpage hourly while the paid version monitors a web page as often as once every 10 minutes. There are also restrictions on how many web pages one can monitor.[14] If a change is detected, Follow That Page! emails the differences in the webpage to the subscriber.

Since we monitored 2 web pages, Follow That Page! alternated checking The Huffington Post and Propublica homepages, or, said another way, each homepage was checked every other hour. Sometimes the changes were significant, for example, one story replacing another on the homepage while at other times the changes were trivial, perhaps a change in punctuation. We did not attempt to differentiate between significant and trivial changes.

During the week June 23-30 Follow That Page! detected 92 changes to The Huffington Post's homepage and 30 changes to Propublica's homepage[15]. This means that nearly every time The Huffington Post's homepage was checked, it had changed. This is not surprising since The Huffington Post is a very active news site[16] with a large number of contributing journalists and editors. But since we did not check The Huffington Post homepage more often, we have no idea how often it actually changed. On the other hand Propublica is targeted to a US audience and has a much smaller staff.

Besides The Huffington Post's or Propublica's internal archives, does any other organization archive their news stories in any way? We looked at the The Internet Archive's

---

[13] Stories no longer visible on the Web may have been deleted or may be archived in a publisher's data centers. If they are deleted, they are gone forever. If the stories are in a publisher's data center, they may or may not be indexed by search engines -- this depends on whether the the publisher's policies and its robots.txt file permits indexing and on whether the archived stories are published behind a paywall or intended only for internal use.

[14] These were the Follow That Page! policies at the time this paper was written. There is no guarantee that these policies will not change.

[15] An arithmetically astute reader will note that if changes are monitored every other hour, at most 7 x 12 = 84 changes could be detected. Follow That Page! reported 92 detected changes for The Huffington Post. This is because page monitoring for the The Huffington Post began at 1.00 and ended at 23.00 while for Propublica page monitoring began at 2.00 and ended 22.00, or (infrequently) at 23.59.

[16] According to the web analytics service Alexa, The Huffington Post has a world pagerank of 158 and a US pagerank of 36.

Waybackmachine[17] crawl history to see what it did. Itshows 2 snapshots of Propublica, one at 14:30:15 and one at 114:30:16 on June 28 -- nothing else.

The Waybackmachine pays more attention to The Huffington Post with 38 snapshots on June 23, 40 on June 24, 36 on June 25, 42 on June 26, 37 on June 27, 42 on June 28, 41 on June 29, and 44 on June 30.

How much attention does the Waybackmachine give to TownNews.com clients? For some idea of this we look at the Waybackmachine crawl history for the 2 TownNews.com newspaper mentioned above, the Coronado Eagle & Journal and the St. Louis Post-Dispatch. Over the same time period (June 23-30) the Waybackmachine crawled the St. Louis Post-Dispatch website twice, once on June 26 and once on June 30; it did not crawl the Coronado Eagle & Journal at all.

No one can fault the Waybackmachine for not giving more attention to these news sites; it's a single donor-funded non-profit organization trying to preserve as much of the visible Internet as possible. The Waybackmachine does what it can, but it alone cannot preserve all born digital news, not even if one considers only the news produced by traditional news publishers.

5. Web-based Archiving and Preservation

Many of the technical issues involved in news preservation are, ultimately, web archiving issues. 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. Recent reports have shown web archiving escalating as a core collection development activity.[18] Though libraries and archives recognize the need for archiving web-published materials, serious technical, institutional, and communal challenges remain.

While the web continues its exponential growth and technical advances accelerate, web archiving program development at research libraries remains underfunded and fractionally staffed.[19] 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

---

[17] For anyone not familiar with the Waybackmachine' here's what it's FAQ says about its purpose:

> *Most societies place importance on preserving artifacts of their culture and heritage. Without such artifacts, civilization has no memory and no mechanism to learn from its successes and failures. Our culture now produces more and more artifacts in digital form. The Archive's mission is to help preserve those artifacts and create an Internet library for researchers, historians, and scholars. ... Much of our archived web data comes from our own crawls ... Internet Archive's crawls tend to find sites that are well linked from other sites. The best way to ensure that we find your web site is to make sure it is included in online directories and that similar/related sites link to you.*

https://archive.org/about/faqs.php#The_Wayback_Machine (Last accessed 2016-07-29).

[18] https://blogs.loc.gov/digitalpreservation/2014/10/results-from-the-2013-ndsa-u-s-web-archiving-survey/
[19] https://blogs.loc.gov/digitalpreservation/2014/10/results-from-the-2013-ndsa-u-s-web-archiving-survey/

complex born-digital acquisitions, including new content.[20] 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. These can only be produced via case studies that enable end-to-end models that libraries and archives can adopt and adapt. Other recognized areas of need include formal training and education models for working with web archives[21] and increasing the currently low uptake among archiving institutions for preserving their web archives locally.[22]

6. Dodging the Memory Hole: Collaboratively Building 21st Century News Collections

The sheer scale of the problem, and the rapidity and diversity of news content today, necessitate community approaches to coordinated collection development. The primary challenge in securing today's news for tomorrow's readers and researchers is establishing relationships between those that produce it and those who have the mission and mechanisms to preserve it.

Hosted by the Reynolds Journalism Institute and the University of Missouri Libraries, the 2014 "Dodging the Memory Hole: Saving Born-digital News Content Forum" (DtMH I, Nov 10-11, Missouri), involved more than 70 representatives from a variety of stakeholder communities in a two-day moderated conversation and planning session about how to jointly develop the capacity to transform born-digital news preservation practices field-wide. The event 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. This highly participatory event brought representatives from these critical stakeholder communities into structured conversation and debate to foster better understandings of each group's perspectives, goals, and barriers to collaborative work.

During DtMH I, participants workshopped on-site a common agenda for change and launched "Take Action!" teams based on the themes that emerged during the event. A total of eight "Take Action!" teams worked on a range of concrete projects in the six months following the event, including an environmental survey in NC, preservation workflows for particular content types, and legal/IP studies.

A second event, "Dodging the Memory Hole II: An Action Assembly" (DtMH II, May 11-12, 2015, North Carolina) was hosted by the Educopia Institute in partnership with RJI and the Charlotte Mecklenburg Public Library in May 2015. Again, the event brought together news publishers, journalists, press associations, distributors, and journalism schools with librarians,

---

[20] For example, in a March 2012 focus group about acquiring born-digital newspapers hosted by the "Chronicles" project lead with five major research libraries who are early adopters in the digital preservation field, four of the five reported that they then considered all born-digital newspaper content to be outside of their collecting scope. They reported that they would need clear legal agreements, ingest and preservation workflows, and justification for its research value in order to begin to convince their administrations that this content is within their missions to collect and curate.

[21] http://library.harvard.edu/03112016-1203/harvard-library-environmental-scan-web-archiving

[22] https://archive-it.org/blog/post/state-of-the-warc-our-digital-preservation-survey-results/

archivists, technologists, and information schools. This second event provided a concrete deadline for the "Take Action!" teams, each of which hosted an action session at the event to share progress and invite collaborative action from the attendees.

At the DtMH II event, attendees identified born-digital news content—particularly that which is both produced and disseminated online—as the key content type in need of concrete workflows, guidelines, and tools for preservation. They also formed a strong informal network (see SaveNews.org and Journalism Digital News Archive) that is raising awareness of the steady disappearance of our nation's born-digital news legacy, and that is cultivating action within and across all of the stakeholder groups involved in this issue.

The upcoming DtMH 2016 event, hosted by RJI, the University of Missouri Libraries, UCLA, and Educopia Institute in Los Angeles, California (October 13-14, 2016) will provide additional opportunities to gather major stakeholders to address this problem. This event focuses on the preservation of news produced not just by mainstream publishers, but by alternative online sources and citizen journalists. It will highlight the increasingly important role of data in the news production cycle, the emergence of new mechanisms for news dissemination within social media environments (e.g., facebook), and encourage stakeholders to take action, locally and nationally, to make incremental progress towards our goal of preserving digital news.

Any strategies for sustaining digital newspaper content must start with catalyzing conversations and connections between these stakeholder communities. By establishing a concrete, shared vision across these stakeholder groups, the DtMH series of events is laying the groundwork for future collaboration and alignment.

7. Summary

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. The case studies, surveys, and professional conferences described in this paper clarify this newly emergent state of affairs in which traditional roles, models, and expectations for long term preservation of news content do not ensure long term survival for new categories of digital news 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. The emerging field of web archiving has begun to address the issues involved in capturing and preserving the ephemeral content of websites. Web archiving is still a nascent field, however, with uncertainty concerning the long term roles, rights, and responsibilities involved in web content curation. More experience with the institutionalization of web archiving will be be needed before it can provide a robust set of models for addressing the the crisis of ephemerality in digital news content.

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.

The fact that a diverse array of participants have begun to come forward in professional events for these purposes is an indication that the larger field has started to acknowledge the importance of this crisis of digital news ephemerality.