

Scanning the Environment: North Carolina Born-Digital News Preservation Practices 2015

Katherine Skinner (Educopia Institute) and Nick Graham (North Carolina Digital Heritage Center)

Introduction

From the Clinton-Lewinsky scandal of 1998, to the Ferguson, Missouri violence of 2015, big news stories now break first online, not in print. These stories are updated continuously as stories shift and morph. They are viewed on mobile devices and experienced through social media. And their presence online is fleeting--a headline stays visible and viable only for a discrete period of time.

Consistently, both research and experience have demonstrated that digital content that is not deliberately organized and managed does not persist over time. In other words, curating data for long-term access requires that content be actively prepared in order to be preserved—simply backing up poorly organized and highly diverse content in broad format ranges does not fulfill the goal of having valid and accessible content in the future.

Lawyers and historians, journalists and economists, politicians and genealogists, and myriad other professions and interest groups all depend upon historical news sources in order to do their work. Now that news is digital-first, where is that “first draft of history” located? Who is acquiring, disseminating, and preserving it? For how long is it stable, and who bears responsibility for its longevity?

Background

What little is known about news preservation practices in the United States is alarming. In 2012, the Pew Research Center’s Project for the Excellence in Journalism cited that already “the emerging world of community online news, less than a decade old, can be difficult to access.”¹ A 2012 survey of news librarians conducted by the Educopia Institute and The Bishoff Group concurred, revealing that most U.S. newspapers are maintaining their digital news files primarily for near-term access (five years or less), and are not yet seeking to ensure the longevity of this record.² A 2014 survey of news publishers conducted by Reynolds Journalism Institute in Missouri similarly demonstrated that most American media enterprises fail to adequately process their born-digital news content for long-term survival.³ And a 2014 survey conducted by

¹ See Pew Research Center:

<http://stateofthemedias.org/2012/mobile-devices-and-news-consumption-some-good-signs-for-journalism/how-community-news-is-faring/?src=prc-section>

² Educopia Institute /The Bishoff Group 2012 Survey on Long-Term Archiving of Born Digital Newspaper Content.

³ McCain, Edward. Reynolds Journalism Institute and University of Missouri Library 2014 Survey on Newspaper Preservation Practices.

the University of Maine showed that while around 50% of the responding news organizations report that they are maintaining content, they largely report doing so by “storing files on servers, in the cloud, and through a third party service”—actions that are unlikely to qualify as “preservation”.

If publishers are not preserving this content, who is? Historically, libraries, historical societies, and archives have collected print and microfilm-based news. But are these organizations collecting born-digital news content?

To answer this question, a team of researchers in North Carolina conducted statewide surveys in 2015 directed at two groups: journalists and librarians/archivists. The results of this investigation are below.

Methodology

The Educopia Institute, in collaboration with the State Library of North Carolina, the State Archives of North Carolina, the University of North Carolina at Chapel Hill School of Journalism and Mass Communications, the North Carolina Digital Heritage Center, Wake Forest University Library, and the Charlotte Mecklenburg Public Library, and Gather Consulting designed two survey instruments in January 2015—one geared towards NC libraries and archives, and another towards NC news editors. These surveys were primarily designed to establish a baseline of current born-digital news preservation activities across the state.

Timeline

- March 10 launch: NC library/archives
- March 14 launch: NC news editors
- April 3 close

Data Collection

Surveys were administered via SurveyMonkey between March 10 and April 3, 2015.

Library and Archives Survey

This survey was designed to collect information regarding acquisition, curation, and preservation of born-digital news content in NC. Survey invitations were sent to individual emails (to all academic and public library directors in the state) and listserv announcements (Society of North Carolina Archivists SNCA listserv). Follow-up emails were sent a week before the close of the survey. Respondents to the survey numbered 103, a sample size that can suggest trends but cannot support statistical generalizations.

News Editor Survey

The survey’s questions sought to collect information regarding the use and management of materials originating in a digital form in North Carolina news organizations. The NC News editor

survey was distributed through a multi-phased approach. Individual letters were mailed to every senior editor of the state's daily and weekly newspapers, as identified through the North Carolina Press Association Directory. Follow-up phonecalls/voicemails were left for all daily editors five days later. Follow-up emails were also issued to senior editors from all daily and weekly newspapers in the state. Only eight news editor responses were obtained.

Multiple choice, open text, and likert scale questions were included in both surveys (see Appendices A and B below). According to inherent question logic, respondents were prompted with up to 15 questions (Libraries and Archives) or 20 questions (News Editors). Within the email invitations to the surveys, and also to the introductory text in the survey instruments, each respondent was advised to budget 10 minutes (Libraries and Archives) or 15 minutes (News Editors) to complete the survey, with the option to return over multiple browser sessions. Once a survey was completed, its responses could not be revised.

Data Analysis

Due to the limited size of the respondent pools, the research team employed qualitative analysis and quantitative, descriptive statistics as its two primary toolsets.

Findings

Born-digital newspaper content quickly and quietly is disappearing, and painfully few are aware of the loss of this crucial historical record.

Print and Microfilm Still Dominate Newspaper Collecting by Libraries

While the types of news media available to libraries are far greater than in the past, print copies and microfilm are still by far the most common formats acquired and preserved. Of the libraries that collect news content, 78% report collecting print newspapers while 61% report collecting microfilm. A little over 20% of the libraries collect digital news content in some format (website files, blog posts, or "digital-first" content).

Most of the Born-Digital News Content Collected in North Carolina is Campus Media

The free-response answers to a question about the types of born-digital news content being collected reveal that nearly all of this kind of work is happening in college and university libraries and is focused on the media produced at their own institutions. Many of the respondents collect PDFs and news photographs from campus newspapers and news services and several reported harvesting campus websites.

Large-Scale Collecting of Born-Digital Community News Content is Not Happening in North Carolina

The results of the survey and related research and conversations have shown that libraries and archives in North Carolina are not actively working to collect and preserve born-digital

community news content. Twenty-two of the 103 respondents report that they are collecting PDFs (some noted they are doing so via CD) and 21 report that they are harvesting news websites. Others are still relying on microfilm (57) and news aggregation by vendors (such as LexisNexis, Newsbank, etc.) as their only means of collecting and providing access to local news content. Those that are collecting born-digital news often are not preserving it--indeed, responses to the question "What percentage of your born-digital news content are you currently preserving" yielded worrisome results: almost 40% of respondents cited preserving 25% or less of their born-digital news content, and only 23% report preserving 100% of this content.

Digital Preservation May Need to be Solved First

In response to the question, "Does your organization have a digital preservation program?," only four respondents said yes. Any responsible effort to collect and preserve born-digital news is going to have to have a reliable and cost-effective digital repository behind it. It's clear that too many libraries in North Carolina are not ready to tackle digital preservation on a meaningful scale.

Money is Not the Only Obstacle to Getting this Done

When asked "Why do you not acquire and preserve more born-digital content?," respondents cited acquisitions difficulties, selection principles, and intellectual property issues as being of equal or greater significance than the cost of managing and storing the content. At this point in the discussion about collecting born-digital news content, the prevailing question is not "How much?" but simply, "How?"

News Publishers Are Not Actively Preserving Born-Digital Content

Although only eight people responded to the news editor survey, their answers clearly demonstrate that their organizations are not actively or officially managing digital content with long-term intentions. None of these respondents report having any written policies for managing born-digital materials. Half of these respondents also indicated that the lack of or insufficient preservation policies and plans were the top threat to this content. None report that they are currently generating revenue by providing access to their older (non-current) born-digital content. These news respondents cite local selection principles and intellectual property/rights issues as the factors that most influence how much born-digital news content their organizations maintain.

Our born-digital news legacy in North Carolina is threatened by a lack of long-term access and preservation intentions, statewide and across stakeholder groups. Addressing this issue swiftly is a critical concern for those charged with the maintenance of the state's historical record.

Project Team

This survey work was undertaken by the following members of the "Dodging the Memory Hole"

Nicholas (Nick) Graham, North Carolina Digital Heritage Center
William (Bill) Kane, Wake Forest University
Caroline (Cal) Shepard, State Library of North Carolina
Katherine Skinner, Educopia Institute
Helen Tibbo, University of North Carolina at Chapel Hill
Ryan Thornburg, University of North Carolina at Chapel Hill
Michelle Underhill, State Library of North Carolina
Rochelle Williams, Gather Consulting
Ashley Yandle, State Archives of North Carolina

It was also supported by the Advisory Board of the [Chronicles II](#) project:

Liz Bishoff, Dorothy Carner, Martin Halbert, Cathy Hartman, Edward McCain, Gail McMillan, Mary Molinaro, Bob Mong, David Thompson, Maria Tucker, Deanna Ulvestad, Rochelle Williams

Appendices

Appendix A: Survey, News Editors

Appendix B: Survey, Libraries and Archives

Appendix A:

Welcome - Born-Digital News Survey

This brief survey is designed to collect information regarding the use and management of materials originating in a digital form in North Carolina news organizations. The survey has 20 questions and will take approximately 15 minutes to complete.

The shift from a print-to-digital news environment has massive implications for the persistence of our news memory. We need your help in identifying what mechanisms news organizations are employing to ensure the longevity of the news records they produce. We want to know how news organizations are backing up their digital news content...whether you are doing that work internally or through external partnerships...and how/whether you plan to provide long-term archiving for this news output for use by future researchers, journalists, and citizens.

We will use the survey results to establish a solid baseline of practice today in NC and to propose future projects that address any gaps or problems we may identify. We will report out on our initial findings at an upcoming NC-based gathering of journalists, librarians, scholars, technologists, publishers, vendors, and other interested stakeholders on May 11-12, 2015, in Charlotte, NC: Dodging the Memory Hole (<http://educopia.org/events/dmh>).

Your participation is voluntary. All the information that you will provide will be kept completely anonymous, so that you cannot be identified.

If you have questions about the survey, please contact Katherine Skinner at (404) 783-2534 or katherine@educopia.org.

For the purpose of the survey, "born-digital" content refers to materials that originate in a digital form, not scanned from other media. Examples include digital photographs, digital documents, digital video, social media files, etc.

Born Digital News Survey

1. Do you have a news librarian or archivist in your news organization?

- Yes
- No
- Don't know

2. Please identify all of the born-digital* content types that your news organization produces. (Please select all that apply.) *For the purpose of the survey, born-digital content refers to materials that originate in a digital form, not scanned from other media. Examples include digital photographs, digital documents, digital video, social media, and etc.

- print-production files
- website files
- digital-first content files (e.g. published online exclusively or prior to a print version) social media files (e.g., blogs, comments, forums, polls)
- mobile-only content (e.g., tablets, smartphones)
- streaming video
- Other (please specify)

3. Please indicate how you store the following types of born-digital* news content files:

CD/DVD | hard drive | external Drive | On-site server | Off-site server | Cloud | Tape | Don't know

print production files

website files

digital-first content files

social media files

mobile-only content files

video files

Other (please specify)

4. How long do you intend to maintain and/or archive these born-digital* news files?

<1 year | 1-2 years | 3-5 years | 6-10 years | >11 years

print production files

website files

digital-first content files

social media files

mobile-only content files

video files

Other (please specify)

5. *What factors influence how much born-digital* news content your organization manages/archives? Please rank from most important (1) to least important (4).*

- *local selection principles*
- *cost of storage*
- *cost of managing content*
- *intellectual property and rights issues*

6. *Where is your born-digital news content regularly backed up?*

- *onsite*
- *offsite in one location*
- *offsite in multiple locations*
- *in a cloud-based environment*
- *we do not regularly back up our news content*
- *don't know*
- *other (please specify)*

7. *Do you provide a copy of your analog content in print or microfilm to another organization for long-term archiving?*

- *yes, to a public library*
- *yes, to an academic library*
- *yes, to a commercial vendor*
- *yes, to another entity*
- *no, we do not provide a copy of our print news to another organization for long-term archiving*
- *please describe*

8. *Do you provide a copy of your born-digital news content to another organization for long-term archiving?*

- *yes, to a public library*
- *yes, to an academic library*
- *yes, to a commercial vendor*
- *yes, to another entity*
- *no, we do not provide a copy of our print news to another organization for long-term archiving*
- *please describe*

9. *If you are not already working with a library or archive to care for the long-term archiving of your born-digital files, would you be interested in doing so?*

- *yes*
- *no*
- *don't know*
- *other (please specify)*

10. Please rank the following concerns you may have about working with libraries or archives to preserve your born-digital news files (from 1 to 5 where 1 is most important and 5 is least important)

- *controlling access*
- *loss of future revenue from content*
- *legal issues*
- *technical issues*
- *time constraints*

11. Have you ever experienced a significant loss of digital news content

- *yes (please describe below)*
- *no*
- *don't know*

12. Do you store and retrieve born-digital news content using a web server or a CMS? If so, please specify what kind

- *yes, locally run*
- *yes, run via our parent organization*
- *yes, run through a vendor*
- *no, we do not use a CMS or web server*

13. Does your organization or the organization you work with to archive your born-digital content have a digital preservation program? Preservation includes backing up data, but requires a more complex series of activities, including managing and migrating formats over time.

- *yes (please describe below)*
- *no*
- *don't know*

14. Who makes decisions about safeguarding born-digital content at your news organization?

- *Publisher*
- *Editor*
- *Technology officer*
- *Board*
- *Other*

15. Does your news organization currently have any written policies for managing born-digital materials?

- *yes*
- *no*
- *don't know*

16. On a scale of 1 to 5 where 1 is very valuable and 5 is not at all valuable, rank the value of long-term access to your newspaper's past coverage

- *for access by your own organization and staff*
- *for access by your audience or community*
- *for producing quality journalism*
- *for producing a legacy for your organization and staff*
- *for producing long-term returns on investment*

17. On a scale of 1 to 5 where 1 is the greatest threat and 5 is not at all a threat, how would you rank the following factors as threats to the loss of born-digital content at your news organization within the next 3 years?

- *physical conditions*
- *technological obsolescence*
- *malicious attacks*
- *lack of or insufficient preservation policies/plans*
- *lack of or insufficient resources for preservation*

18. Are you currently generating revenue by providing access to your older (non-current) born-digital content? If so, please specify how and what vendor (if any) you use.

- *yes*
- *no*
- *don't know*

19. In what other ways are you generating revenue from your non-current digital news assets?

20. Would you like to be contacted about future work around born-digital news preservation issues?

- *yes*
- *no*
- *other*

Appendix B:

Welcome - Born-Digital News Preservation Survey

This brief survey is designed to collect information regarding the acquisition, use, and management of born-digital news materials in North Carolina. The survey includes 15 questions and will take approximately 10 minutes to complete.

The results of the survey will help us better understand how libraries and archives like yours interact with news organizations and acquire news content in today's new media landscape.

Your participation is voluntary. All the information that you will provide will be kept completely anonymous, so that you cannot be identified.

If you have questions about the survey, please contact Katherine Skinner at (404) 783-2534 or katherine@educopia.org.

For the purpose of the survey, born-digital content refers to materials that originate in a digital form, not scanned from other media. Examples include digital photographs, digital documents, harvested web content, digital manuscripts, electronic records, and etc.

Born-Digital News Preservation Survey

1. What type of organization do you represent?

- library-public
- library-academic
- library-special
- archives-government
- archives-academic
- archives-corporate
- archives-religious
- historical society

Other (please specify)

2. What forms of news content does your library/archives currently maintain? (select all that apply)

- print newspapers
- microfilm/microfiche
- historical video/TV
- print-production files
- website files
- digital-first content files (e.g. published online exclusively or prior to a print version)
- social media files (e.g., blogs, comments, forums, polls)
- mobile-only content (e.g., tablets, smartphones)
- streaming video

Other (please specify)

3. Does your library/archives currently acquire and maintain born-digital* news content?

- Yes
- No (skip to Question 15)
- Don't Know

*For the purpose of the survey, **born-digital content** refers to materials that originate in a digital form, not scanned from other media. Examples include digital photographs, digital documents, harvested web content, digital manuscripts, electronic records, and etc.

Born-Digital News Preservation Survey

4. Please briefly describe the born-digital* (NOT digitized) news collections that you manage (e.g., how many publications, how much content, what formats?).

5. What rights does your library/archive have for archiving and displaying the born-digital* news content you acquire? (Select all that apply.)

- Right to display the news content
- Right to copy the news content
- Right to archive the news content
- Right to preserve the news content
- Don't know

Other (please specify)

6. Does your organization maintain relationships with any of the following groups around born-digital* news acquisition? (Select all that apply)

- newspaper
- press association
- media outlet
- vendor
- none of the above

Other (please specify)

7. Do you have standard agreements for donations/acquisitions of born-digital* news content?

Yes (please specify type below)

No

Don't Know

Please specify:

*For the purpose of the survey, born-digital content refers to materials that originate in a digital form, not scanned from other media. Examples include digital photographs, digital documents, harvested web content, digital manuscripts, electronic records, and etc.

Born-Digital News Preservation Survey

8. Please indicate how you receive the following types of born-digital* news content files:

	CD/DVD	External Drive	Server-based upload (e.g., FTP)	Cloud	Web harvest	Don't know
print production files	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
website files	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
digital-first content files	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
social media files	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
mobile-only content files	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
video files	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Other (please specify)

9. How long do you intend to maintain and/or archive these born-digital* news files?

	<1 year	1-2 years	2-5 years	5-10 years	>10 years
print production files	<input type="radio"/>				
website files	<input type="radio"/>				
digital-first content files	<input type="radio"/>				
social media files	<input type="radio"/>				
mobile-only content files	<input type="radio"/>				
video files	<input type="radio"/>				

10. Where is your born-digital* news content regularly backed up?

- onsite
- offsite in one location
- offsite in multiple locations
- in a cloud-based environment
- we do not regularly back up our news content

Other (please specify)

11. Does your organization have a digital preservation program? Preservation includes backing up data, but requires a more complex series of activities, including monitoring and migrating formats.

Yes (please describe below)

No

Don't Know

Describe your digital preservation program:

*For the purpose of the survey, born-digital content refers to materials that originate in a digital form, not scanned from other media. Examples include digital photographs, digital documents, harvested web content, digital manuscripts, electronic records, and etc.

Born-Digital News Preservation Survey

12. What percentage of your born-digital* news content are you currently preserving?

- None
- 1-10%
- 11-25%
- 26-50%
- 51-75%
- 76-99%
- 100%
- Don't know

13. Why do you not acquire and preserve more born-digital* news content? Please rank from most important (1) to least important (5).

<input type="text"/>	acquisition difficulties
<input type="text"/>	local selection principles
<input type="text"/>	cost of storage
<input type="text"/>	cost of managing content
<input type="text"/>	intellectual property and rights issues

14. Have you ever experienced a significant loss of digital news content?

- Yes (please describe below)
- No
- Don't Know
- Please describe

*For the purpose of the survey, born-digital content refers to materials that originate in a digital form, not scanned from other media. Examples include digital photographs, digital documents, harvested web content, digital manuscripts, electronic records, and etc.

Born-Digital News Preservation Survey

15. Would you like to be contacted about future work around born-digital news preservation issues?

- Yes
- No
- Other (please specify)

16. Please provide us with the following demographic information. We will not contact you unless you have indicated that you would like further information about born-digital news preservation.

Name

Organization Name

Job Title

City/Town

State/Province

ZIP/Postal Code

Country

Email Address

Phone Number

Born-Digital News Preservation Survey

Thank you so much for taking this survey on Born-Digital News Preservation. The results of the survey will help us improve the ability of libraries and archives to acquire and preserve news content in today's multimedia landscape. If you have questions about this survey, please contact Katherine Skinner (katherine@educopia.org).

If this survey has peaked your interest in this topic, please consider joining librarians, archivists, technologists, journalists, publishers, and researchers for the upcoming "Dodging the Memory Hole" Action Assembly on preserving born-digital news in Charlotte, NC on May 11-12, 2015 (<http://educopia.org/events/dmh>).