

# News Archive

CHAOS

Kathleen A. Hansen

Nora Paul

University of Minnesota

School of Journalism and Mass Communication

## **News Archive Chaos: A Case Study**

**Kathleen A. Hansen and Nora Paul**  
School of Journalism and Mass Communication  
University of Minnesota  
Minneapolis, Minnesota, USA  
[k-hans@umn.edu](mailto:k-hans@umn.edu); [npaul@umn.edu](mailto:npaul@umn.edu)

Paper to be presented at the IFLA News Media and Audiovisual and Multimedia Sections' Conference "Transformation of the online news media: implications for preservation and access", Stockholm, Sweden, National Library of Sweden, 15-16 April 2015



Summer 2015

# Newspaper Research Journal

Reading Today's News  
on Future Computers:  
The particular challenges  
of archiving interactive  
news applications

Preserving Newspaper  
Information Assets:  
Current practices and  
ongoing concerns

Preserving Ethnic  
News Media in the  
Digital Environment

News Preservation  
Through Partnerships

JDNA: An Agenda  
to Save Born  
Digital News  
Content

Preserving News as  
History: Perspectives  
on born digital news in  
the Pine Tree State

What? No Backups?:  
Preserving hard copy  
newspapers in the digital  
age

<b>Archiving Publications</b> <i>Preserving the entire published news report</i>	<b>Archiving Stories</b> <i>Preserving the individual story packages</i>	<b>Archiving Components</b> <i>Preserving specific file types used in stories</i>	<b>News Functions</b> <i>(news materials that fulfill those functions)</i>	<b>Archiving Issues</b> <i>Obstacles to the preservation of materials</i>	<b>Stakeholders</b> <i>Who needs to care about preservation</i>
Newsprint -- Current -- <u>Backfiles</u> (bound vol., microfilm)	News stories -- Current -- <u>Backfiles</u> (clippings, databases)	Text	Legal record ( <i>public notices</i> )	Backward compatibility	Archivists- Internal to news organization
Web Publication version	Web stories (includes related <u>mat'l.</u> )	Photos	Historical record ( <i>overall package, specific stories</i> )	Forward compatibility	Archivists - External to news organization
Mobile Application version	Mobile stories (includes photo, videos)	Video	Record of commerce ( <i>advertising / classified</i> )	Transition to different publishing platforms	Reporters / editors / researchers
Social Media Sites	Posted stories (includes comments, contributed content)	Audio	Record of public affairs in a community ( <i>calendars / public notices</i> )	Ceased publications (who maintains past files?)	Computer programmers / coders
		Charts / Graphics / Maps	Genealogical ( <i>obituaries / paid and non-paid</i> )	Access / pricing models (free, freemium, premium)	Historians / Genealogists
		News apps (programming to create interactive)	Community forum (letters to <u>editor, comments</u> )	Born digital (and belief that it will stay)	Content developers / business strategists
		Data sets		Sale of archives to for-profit vendors	Lawyers
		Links (embedded or sidebar)		Donation of archives to non-profit orgs.	Commercial data vendors
		Corrections		Disinterest / lack of appreciation for value	Citizens
		Comments (on web version)			<b>Government agencies</b>

# Guidelines for Digital Newspaper Preservation Readiness

## Katherine Skinner & Matt Schultz

### March 2014

¶ 1 Identifying what digital newspaper content an institution possesses is the first step in understanding its current and future preservation needs. Digital newspapers are created and acquired by institutions under a diverse array of circumstances, over wide spans of time, and often under the care of multiple managers. For that reason, an institution's digital newspaper collections and corresponding content files may reside on a range of different storage media (external hard drives, CDs, disk arrays, tape systems, etc.) and in multiple locations.



¶ 5 A basic inventory may provide a description of an institution's full range of newspaper files, including title, content type, format, size, associations (title/issue and page/article information), and file location. A more detailed inventory might also include information about creation date, copyright/permissions, software needed to render objects, checksums, and digital object identifiers.



## CONTENT

- Entire publication - newsprint
- Individual stories – newsprint
- Entire publication – website
- Individual stories - website

## LOCATION

- Internal archive
- External collections
  - Public libraries
  - University libraries
  - “Memory institutions”
- Commercial collections

## FORMAT

- Hard copy of publication: bound volumes
- Hard copy of stories: clippings
- Microfilm of publication: microfilm
- Digital version of stories: electronic archive
- Web version of publication
- Web version of stories
- PDF version of hard copy

## FINDING TOOLS

- Clipping files
- Indexes
- Commercial databases



## NEWSPRINT PRODUCT RESEARCH

Question Type	Resource Need
<i>What was the coverage of a specific person or event in a specific newspaper? You know the date</i>	<i>Microfilm Database</i>
<i>Did a particular newspaper have any coverage about a specific person or event? You don't know the date</i>	<i>Scanned microfilm Database Index</i>
<i>How was a specific person or event covered over a range of publications?</i>	<i>Scanned microfilm Database</i>
<i>Content analysis of a news publication for a specific date or range of dates</i>	<i>Microfilm PDF</i>
<i>Analysis of the "commercial information" (advertising, classified ads) for a specific date or range of dates</i>	<i>Microfilm PDF</i>
<i>Determine the relevance of how a news story was "played" in the publication and see all the material in the story package.</i>	<i>Microfilm Scanned microfilm PDF</i>

DULUTH NEWS TRIBUNE ACCESS

	1870s	1880s	1890s	1900s	1910s	1920s	1930s	1940s	1950s	1960s	1970s	1980s	1990s	2000s	2010s	Today
Story/HC					p	p	p	p	p n	p n	p n	p n	p n	p	p	p
Story/DB													de dn	de dh dn	de dn	de dn
Finding Aids				p	p	p					p	p	p			
Pub/HC			h													n p
Pub/MF	#	#	#	#	#	#	#	#	#	#	#	#	#	#	#	#
Pub/SMF		dn	dn	dn	dn	dn										
Pub/PDF																n
Web/In															n	n
Web/Ex														w	w	w

Story/HC= Hard copy of news stories (clips)

Story/DB=Individual stories in searchable text database (ie: Nexis)

Story/SMF=Individual stories located on searchable scanned microfilm

Pub/HC=Entire publication in hard copy (newsprint edition)

Pub/MF=Entire publication on microfilm

Pub/SMF=Entire publication on searchable scanned microfilm

Web/In=Web publication's internal searchable story archive

Web/Ex=External web search engine or archive with publications' material

LOCATIONS

h=Historical Society

n=newsroom

p=public / university library = sporadic and selected story clips / vertical files

dn=database/newsbank

dg=database/gale

de=database/Ebscohost

dh=database/Highbeam

mf=proquest

w=web (ie: google or wayback machine)

#=available at a number of locations (but not all locations have full run)





# StarTribune




# Newspaper extinction timeline

When newspapers in their current form will become insignificant\*



USA	2017
UK, Iceland	2019
Canada, Norway	2020
Finland, Singapore, Greenland	2021
Australia, Hong Kong	2022
Denmark	2023
New Zealand, Spain, Czech, Taiwan	2024
Poland, Sweden, Switzerland	2025
South Korea, Metro Russia, Belgium	2026
Netherlands, Ireland, Metro Brazil, Italy	2027
Austria, Slovakia, Greece, Portugal, U.A.E.	2028
France, Israel, Malaysia, Croatia	2029
Germany, Estonia	2030
Japan, Metro China	2031
Hungary, Lithuania	2032
Latvia, Metro Mexico	2033
Serbia, Saudi Arabia	2034
Bulgaria, Chile, Uruguay	2035
Russia, Turkey	2036
Metro South Africa, Thailand	2037
Mongolia	2038
Argentina	2039
Rest of the world	2040+

 – stripes indicate that newspapers will be extinct in metropolitan areas before regional areas

## DIGITAL CONTENT

- Entire publication
- Individual stories
- Video
- Audio
- Interactive Graphics
- Comments

## LOCATION

- Internal archive
- Web archives (Google news? Wayback Machine?)
- Commercial collections

## FORMAT

- Web version
- Tablet version
- Mobile version
- Social media version
- Future delivery / appliances?

## FINDING TOOLS

- Search engines
- Commercial databases?

## WEBSITE PRODUCT RESEARCH

### Question Type

### Resource Need

*What was the coverage of a specific person or event from a specific news organization?*

*News website search  
External website search*

*How was a specific person or event covered over a range of publications?*

*External website search (Google)  
(if the publication allows)*

*Content analysis of a news publication for a specific date or range of dates*

*???*

*Analysis of the “commercial information” (advertising, classified ads) for a specific date or range of dates*

*xxx  
(different people get served different ads)*

*Determine the relevance of how a news story was “played” in the publication, see all the material in the story package.*

*???*

*See all the material related to a specific story (comments, photos, video, interactive graphics)*

*Maybe / maybe not*

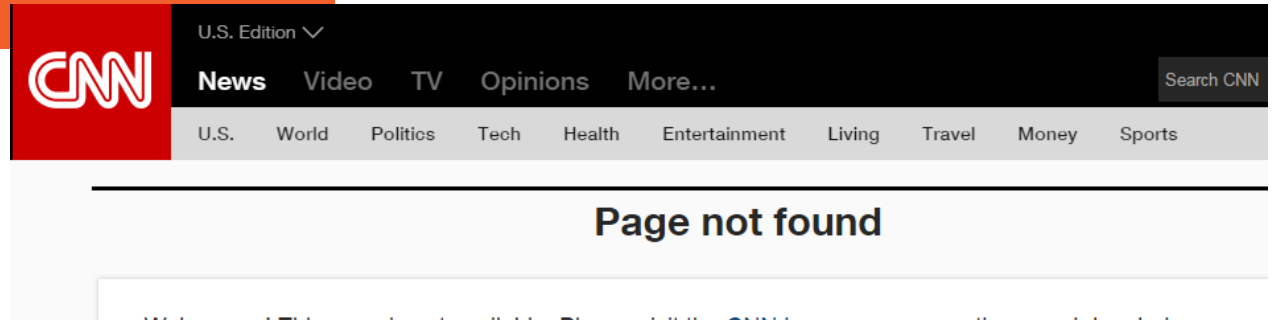
# Page not found

The requested page "/story.html?StoryID=15093" could not be found.

**ALTERNET**

# Page not found

The requested page "/views03/0131-01.htm" could not be found.



The screenshot shows the top navigation bar of the CNN website. On the left is the CNN logo. To its right are links for 'U.S. Edition', 'News', 'Video', 'TV', 'Opinions', and 'More...'. Below these are category links: 'U.S.', 'World', 'Politics', 'Tech', 'Health', 'Entertainment', 'Living', 'Travel', 'Money', and 'Sports'. A search bar labeled 'Search CNN' is on the far right. The main content area displays the text 'Page not found' in a large, bold font.

# Forbidden

You don't have permission to access /voting\_machines.htm on this server.

Additionally, a 403 Forbidden error was encountered while trying to use an ErrorDocument to handle the request.

# The Washington Post

We are unable to locate the page you requested.  
The page may have moved or may no longer be available

You may also want to try our search  
to locate news and information on [washingtonpost.com](http://washingtonpost.com)



This webpage is not available

Reload



**“All great  
changes are  
preceded by  
chaos.”**

**Deepak Chopra**

happytoinspire