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; npaul@umn.edu

News Archive Chaos: A Case Study

Kathleen A. Hansen and Nora Paul

*In spite of its name, the chief function of the newspaper is not to give the news, it is not even exclusively to reflect public opinion – important as this is, -- but it is to record all contemporaneous human interests, activities, and conditions and thus **to serve the future***. [emphasis added]

Historian Lucy Maynard Salmon, 1923

News professionals have long understood their role in capturing both the daily occurrences of a community and a snapshot of contemporary life. In both the "official" content of the newspaper (legal notices, public records) and the unofficial content, including accounts of events, advertisements, editorials, cartoons, and related material, news organizations paint a picture of a particular time and place.

Early "Morgues"

Storing, indexing and making accessible the backfiles of a daily newspaper has never been a trivial task. Typically, the earliest U.S. newspaper morgues consisted of reference book collections and were organized by journalists themselves rather than by professional librarians with information preservation and access training. The *New York Herald* started building a library of books for the news staff in 1845, just 10 years after the paper's founding. Clippings of the day's news content, however, were not systematically saved until 1860. That year the *Herald* also began producing an index to the newspaper content (Semonche, 1993a, p. 5).

By the mid-nineteenth century, however, large newspaper organizations began clipping their own content, spurred by competition among papers for subscribers who wanted accurate, fair and authoritative information (Semonche, 1993a, p. 4). Changes

in news content itself, including a move away from simple reporting and towards more analytical writing, made it necessary for journalists to gain access to background information and specialized subject knowledge as well as their own previous writing on the news pages.

Just as news organizations began collecting and storing their own content, public and academic libraries began building collections of newspaper files for historical reference and community documentation. Members of the American Library Association suggested developing a national newspaper index as early as 1893 (Semonche, 1993b, p. 375). As newspapers started, stopped, merged or were sold, clipping collections made their way into academic and public library collections. Librarians in the U.S. began referring to the *New York Times* as a newspaper of record soon after the newspaper began publishing a quarterly index and marketing it to libraries in 1913 (Rothman, 1966).

In addition to the clip files kept by news organization libraries, most newspapers bound their own issues in order to have a complete record of their output. Academic and public libraries and historical societies also collected and bound newspaper issues, or inherited those bound volumes as news organizations closed, merged or changed hands. Bound volumes of newsprint posed their own storage, preservation and access problems. The highly acidic paper used for newsprint after 1870 resulted in deteriorating, worn, torn, crumbling and faded pages. Volumes would get lost or misfiled. Even when a set of bound volumes could be preserved, there was rarely a comprehensive index to the paper's contents so it was impossible to retrieve a specific article from the shelves of bound newsprint.

Microfilm was once declared the savior of newspaper preservation. *New York Herald* news librarian David G. Rogers worked with Eastman Kodak in 1932 to adapt a microfilm camera to film newspaper files (Mills, 1981) and by 1946 the Bell & Howell Company made the filming of newspapers a major part of its business.   But microfilm poses its own preservation problems. Acetate-based film, which was used up until the 1980s, deteriorates when not stored at the proper humidity and temperature, resulting in the loss of information captured on the film. In most cases, the original issues from which the acetate microfilm was made were discarded. Unless the acetate-based film was checked and the deterioration was discovered in enough time to transfer the images to polyester-based film, the content was lost. Even the most recent microfilm processes cannot prevent scratches, tears and unreadable sections of film when the rolls are not handled and stored properly. And microfilm runs of newspaper content have *never* captured much of the original content such as fold-outs, color reproductions, ad inserts, magazines, color comics, and related material that would be valuable to future researchers.

Bound backfiles and microfilm of yesterday's newspaper did not replace the need for news organizations to continue to clip and file each issue and keep an internal morgue for journalists' use. Most large news organizations in the U.S. continued to build and maintain their clip files until the mid 1980s (or even later) when large-scale adoption of digital archiving systems became the norm. At that point, the clip files became a storage and maintenance nuisance and expense. Most news organizations did not have the budget to microfilm or (later) digitally scan the hundreds of thousands of newsprint clippings stored in envelopes filed in metal filing cabinets. Many

newspapers moved the filing cabinets to a basement storage room or even off-site. Some gave away the clipping files to local historical societies or libraries, which were equally under-resourced to catalog and preserve them. And some organizations simply threw them all away.

News Library Automation Chronology

Starting in the late 1960s in the U.S., the American Newspaper Publishers Association worked with the Massachusetts Institute of Technology to develop a way to capture and make accessible the text produced by the news organizations' electronic typesetting systems that replaced the old "hot type" production process (Metcalf, 1993, p. 220). By the early 1970s, several news organizations were able to offer databases of news article abstracts (*New York Times Infobank* in 1974 and Info-Ky News Retrieval System from the Louisville *Courier-Journal* and the *Louisville Times* in 1976). The first true, full-text retrieval system called Info-Globe was introduced by the Toronto *Globe and Mail* in 1976, followed in quick order by systems such as VU/TEXT, Datatimes, NEXIS and databases offered by online vendor DIALOG. By 1992, more than 150 U.S. newspapers were offering their backfiles through an online system (Metcalf, 1993, p. 230).

The limitations of the systems were many. The text-only databases did not capture placement or play of the stories; images or graphics that accompanied an article were simply mentioned in a note (users had to go to the microfilm or hard copy to see the visuals); agate content (death notices, stock prices, box scores) was usually stripped out; and individual vendors had a variety of restrictions about the inclusion of wire service copy, copy generated by freelancers, different versions of stories produced

for different zoned editions, and so forth. What's more, the internal version of the database that was accessible exclusively to the news organization itself was the *only* truly complete version of what had appeared in the printed newspaper.

Nonetheless, these online, full-text retrieval systems became essential to both the news organizations themselves and to the memory institutions that purchased licenses for access to the content. News library staffs expanded to include workers hired specifically for "enhancing," that is, cleaning up the text files from the front-end production system, adding keywords or descriptors, and ensuring that the content electronically shipped to the vendor on a regular schedule met all the licensing requirements.

News organizations also started to see the potential for generating income from these backfiles. There was a market for old news. The content was valuable not just to the journalists but to the general public. Up to this point, the internal news library staff made decisions about what to keep and make accessible to the journalists inside the organization. For the general public, old news content had typically been collected, bound, microfilmed, indexed and otherwise made accessible through a series of haphazard, idiosyncratic, local decisions made by understaffed libraries or memory institutions, genealogists, enthusiasts or by no one at all. Now news librarians started inking contracts with proprietary vendors who would control what was included and excluded from the public database, how it was priced, who would have access and a myriad of other decisions the consequences of which are felt to this very day.

Newspaper database vendors continued to improve their systems throughout the 1990s, even while industry consolidation reduced the number of viable competitors to

just a handful. Meantime, a number of companies were springing up offering online

content to the public at home. CompuServe, Prodigy, AOL and a host of smaller,

regional companies started providing dial-up access to a variety of types of digital

content - email, bulletin boards, discussion groups and yes, even some news content.

However, the offerings were limited and the technology infrastructure (bandwidth,

modem speeds, etc.) made it cumbersome and difficult to use such systems for

anything except the capture and sharing of brief content snippets.

This was the state of affairs until the mid 1990s when the first news *websites*

began appearing. The San Jose (Calif.) *Mercury News* became one of the first daily

newspaper in the U.S. to launch a true web-based version of its content in January of

1995, the Mercury Center Web (Carlson, 1999-2009). Built on the infrastructure of the

World Wide Web and taking advantage of increasing bandwidth and the public's access

to more robust computing technology, these early news sites mostly consisted of

shoveling the content from the print product onto the screen. However, there was no

thought put into whether and how to archive the content from this new delivery platform.

The Internet Archive's Wayback Machine captured a few pages of Mercury Center Web

starting in 1998 but as with all Wayback captures, those are static screenshots, not a

true archive of the site. It is a tragedy of monumental proportions that the vast majority

of news websites did not capture, and now have no record whatsoever, of what they

looked like and what they offered on the day they launched or, in most cases, for years

afterwards.

Of course, current news websites consist of much more than the content from the

day's print publication. For the "oldsters" among the group, there is now a 20-year

history of publishing on the web, but no corresponding archive of what those 20 years included. And the born-digital sites (those with no print counterpart at all) continue to proliferate and push the boundaries of what is possible using this interactive platform. User-generated content, commenting functions, infographics, multimedia, dynamic databases that can be manipulated by the user and much more are regular fare. And once again, there is no current strategy or system for archiving this content.

Format and Location Challenges

Communications scholar Sandra Gabriele lays out an elegant argument about the material transfiguration of a newspaper from paper to microfilm to database in a 2014 paper. She observes that, "As a cultural form, newspapers - at least since the 1880s - have imposed an order and logic that readers have come to understand and adopt." (Gabriel, 2014, p. 6) Our "modern" conception of a newspaper includes bold headlines spanning multiple columns, discrete news sections (business, sports, society, etc.), color inserts and supplements (especially on Sunday) and recurring features. As print newspaper backfiles were converted to microfilm, they retained the linear, chronological flow of the paper original. Whether accessed through a digital or analog microfilm reader, the logic of reading the newspaper by scrolling through static pages of content remained.

Databases, on the other hand, operate entirely outside of this linear, chronological order. Their organizing principle, with the ability to retrieve and filter data in multiple ways, "bears no resemblance to the ordering schemes imposed by publishers before or after binding, librarians or microfilm companies...Networked databases disrupt the temporal and spatial arrangements that once dominated how one

read a newspaper (on paper or microfilm), making local small-town papers as available - and potentially interesting as research objects - as major metropolitan papers." (Gabriele, 2014, p.10) However, the database destroys the historical order of the newspaper narrative. It also facilitates the creation of enormous "data sets" comprised of news content from across large geographic regions and time periods. And many of these databases are increasingly accessed by the public through libraries (university, public, historical society, etc.), which enter into complex, expensive and highly restricted licensing agreements with a relatively small number of corporations who are making the decisions about what to offer.

The implications of Gabriele's argument are plain to see if one tries to examine the archiving history of any one newspaper. The accretions of preservation decisions across many decades have direct impact on what is available, in which medium, at what cost. Even for a country's largest, most important and prominent newspapers, there are likely to be gaps in the archival record because of these transfigural changes over time. Brought down to the level of an individual community's news history, the archival record quickly devolves into chaos.

In their "Guidelines for Digital Newspaper Preservation Readiness," Katherine Skinner and Matt Schultz (2014) lay out the steps and processes for memory institutions to manage their *digital newspaper content.* The document does not address the management of print or non-scanned microfilm content, but we can take some insights from the Guidelines nonetheless. The first step in the process is to identify the amount, type and location of an institution's newspaper collections.

Producing an inventory requires creating categories of content. For any given newspaper, we might propose that these **content categories** could be useful:

-- a record of the entire publication as it appeared in full - this could include bound copies of the print product, microfilm (analog or digital) of the print product, or a digital facsimile of the print product (PDF files, etc.), and assumes capture of both news and advertising content

-- a record of individual stories as they appeared - this could include clippings file from the print product, microfilm that has been indexed by story or can be searched digitally, searchable digital facsimiles of the print product, or digital news databases that provide story-level access to the portion of the print product that was sent to the commercial vendor or made accessible by the news organization itself using its digital backfiles CMS

-- a record of the *website* content - this could encompass some type of capture of the entire website, the content of individual stories that appeared both in print and on the website, the content of individual features that appeared *only* on the website, news OR advertising content vs. news AND advertising content

Another way to categorize the inventory would be to do so by **format categories** of the archive:

-- hard copy of the publication in its original form - bound volumes

-- hard copy of individual stories -- clippings files

-- microfilm of the publication - analog or digital

-- digital version of the hard copy stories - original electronic archive

-- web version

-- PDF version of the hard copy

Referring back to Gabriele's work, the inventory of types and formats of content carries with it major implications for what the user will actually be able to do with the content if it can be found. Which version will be best suited to the work that the researcher, historian, genealogist or journalist must complete? Is it important for the user to reconstruct the physical object, to simply gain access to the content bereft of form, to be able to understand how content morphed from analog to digital, something else? The simple act of compiling an inventory starts to illuminate what will and will not be possible going forward.

The next step in the inventory is to identify **locations** where these categories of content can be found:

-- Internal archive - What is available to journalists themselves? In which format? For what time period? With what sort of requirement for assistance to gain access?

-- External collections (public libraries, university libraries, national collections, historical societies, enthusiast collections). What is available to the public? In which format? For what time period? At what cost to gain access?

And finally, there is the issue of inventorying the **finding tools** that allow users to gain access to the collections:

-- clipping files of what appeared in the hard copy of the newspaper - created by the news organization and organized by topics - generally available just to the newsroom but sometimes "inherited" by a memory institution

-- indexes to the printed version of the newspaper - typically includes just *story* level access, no agate, no advertising, in many cases no wire service content - indexes might exist as card files, printed volumes, some type of user-generated database stored on a local server, etc.

-- proprietary databases of what appeared in print that was digitally captured and sent to the vendor - again, typically *story* level access but with many restrictions on what was included (agate, wire service content stripped out; no ads, etc.)

-- Web search engines - what appeared in some digital form that can be spidered by the company's search system - might include stories, images, ads, interactives, homepages, etc.

-- Wayback Machine - what appeared on the website (both print and digital-only content) - attempts to capture placement and play of at least the homepage - must recognize that some news organizations (e.g. Gannett in the U.S.) block the Wayback Machine's spiders

The categories above represent just one of many different organization schemes that might be applied to a given community's newspaper history, and different memory institutions in a community could certainly apply different schemes to the exact same content. All of this suggests that a simple directive to "inventory your collection" is not

so simple after all. As the following case study will demonstrate, this may be the greatest challenge facing any institution that wants to start a newspaper preservation project of any sort, not just a digital news preservation project.

Case Study: Duluth, Minnesota

In order to illustrate the availability of the full record of news stories, photos, and online-only content from news organizations the authors decided to detail the archiving options for one mid-sized Midwest city in the United States.

Duluth, located on Lake Superior in northeastern Minnesota, is an international port city with a long history as a major shipping hub for goods of all sorts traveling through the Great Lakes and the St. Lawrence Seaway to the Atlantic Ocean and the wider world.[1] Today, the Duluth News Tribune is the major newspaper for the city. The Duluth Tribune began publication on May 3, 1870 as a weekly and began its run as a daily in 1881. It merged with another regional paper, the Duluth News, in 1892, and was renamed the Duluth News-Tribune. A morning paper, it was purchased by the evening Duluth Herald in 1929 and Duluth, like most mid-to-large sized cities, had two strong newspapers.

In 1936 the morning News-Tribune and evening Herald were both purchased by Ridder Publications (which later merged with Knight Newspapers to become the Knight-Ridder newspaper group.) The ability to support two daily newspapers waned in the

---

[1] Duluth, Minnesota and Superior, Wisconsin (connected by a bridge across a narrow place in the harbor) are sometimes referred to as the Twin Ports. Some Duluth newspaper content is of interest to residents of Wisconsin, hence the Wisconsin Historical Society's collections are mentioned in this case study as well.

1980s in many cities across the United States, and in Duluth the two newspapers were merged, becoming the Duluth News-Tribune and Herald in 1982. The name was shortened to the Duluth News-Tribune in 1988 and in 2000, the hyphen was dropped.

In 2006 the McClatchy Company purchased all of Knight-Ridder's assets, including the Duluth News Tribune but quickly divested themselves of several of the properties they bought, selling the News Tribune to Forum Communications, which has been its corporate owners since June 2006.

The News Tribune had a news library that, according to the assistant to the editor's recollection, started in about 1955 (Gajda, 2015). There were several staff members dedicated to clipping news stories for reference purposes for the newsroom. By 1995, when the newspaper "went digital" and began to store its content in an online archive, the news library was down to one staff member. When he retired, the archive operation was "automated" and the News Tribune library, as with so many others over the past 15 years, shut its doors.

The clipping collection, compiled over forty years of news library operation, has been moved to what is referred to in the newsroom as "the attic." A staff-written blog called "News Tribune Attic" describes the resting place for the old news library collection this way:

> *On the third floor of the News Tribune building, in a back room with warped wooden floors, glass-block windows and just a bit of dust, sit rows of file cabinets stuffed with photos, clippings and rolls of microfilm documenting the history of Duluth, Superior, the Northland and beyond.* (Krueger, 2014)

The categorized clippings and photo prints are available only to the newsroom and success at finding information when needed is by no means guaranteed.

About the time the archiving of newspaper stories went digital, the website of the news product began. The domain name "duluthnews.com" was registered on July 1, 1996 and the earliest capture of the website by the Wayback Machine was on Dec. 9, 1997. In keeping with the "cyberspace" sentiment of the time, the website was initially called "TribUniverse."

The website then, as is largely still the case, contained content duplicative of the print newspaper, with some notable exceptions. The newspaper publishes regional and national news in addition to the local news stories but only the local news can be found on the website. Video content and blogs, special to the website, are some of the augmented materials online that, obviously, cannot be found in the newspaper.

Several years ago, Forum Communications contracted with Olive Software Inc. to provide a PDF version of the News Tribune newspaper edition, essentially coming full circle by providing a digital reproduction of the newsprint news product.

Customers interested in reading the daily news from Duluth's major newspaper have a variety of choices - subscribe to the daily newsprint product delivered to their home, read the news on their computer or mobile device as published on the news website, or subscribe to the PDF version and "flip" through the newspaper on their computer, phone or tablet.

Researchers interested in locating information previously published in the Duluth News Tribune also have a variety of choices - therein lies the "chaos" referred to in the

title of this paper. Depending on the information task at hand, the date the required information was published, and the desired format, researchers may have an easy time locating the news article they need or they may be facing a frustrating, possibly fruitless, task.

In the following section we will provide a picture of the range of materials from the Duluth News Tribune that can satisfy researchers' needs. The patchwork pattern of dates, formats, and finding aids represent the convoluted archiving and availability of one mid-sized newspaper - a pattern that can be seen in newspapers and communities across the country and, in all likelihood, around the world.

HARD COPY: Bound Volumes: Many newsrooms maintained huge "books" of each day's publication. It was also a common practice for newspapers to send daily publications to the U.S. Library of Congress. The printed newspaper was also packaged and sent to the company contracted to provide microfilm. As newsroom costs rose, including the cost of office space, many newsrooms eliminated the bound volumes. The News Tribune was one of them.

With some digging, researchers can find "stashes" of bound volumes in various locations. The periodical archives librarian at the Wisconsin Historical Society said that they have bound volumes of the Duluth newspaper from April 1, 1897 through December 31, 1899, and, for some reason, two issues in 1906 (Larson, 2015).

Clips: The collection of news story clippings from the Duluth News Tribune in-house library dates back to sometime in the 1950s. But unless you work in the newsroom, these files are unavailable.

The University of Minnesota - Duluth has a "vertical file" that contains clippings by subject from various Duluth newspapers. Started by the St. Louis County Historical Society in 1922, it contains story clips up until the mid-1970s (no one is sure when.) In the late 1980s (again, no specific date) the collection was re-started but the effort to clip news stories stopped around 2010.   As the note about the clip collection states, there is little assurance that a topic is comprehensively captured in these clippings: *"These files contain printed material, primarily newspaper clippings. The amount of material and the date range varies considerably by topic. Please contact Archives staff to inquire about content of and access to files."*

The Duluth Public Library has several collections of news stories related to specific topics. For example, there are three scrapbooks containing newspaper clippings from 1959 to 1962 on Duluth politics, housing, employment, shipping, and education that a former Mayor of Duluth compiled.

There are also "hanging files" on Duluth specific topics that librarians continue to update with local publication clippings, including from the newspaper, that have material from the early 1900s.

But there is no source for those seeking a collection of news stories related to a specific topic or person over the span of the newspaper's publication.

Finding Aids: Between 1922 when availability of scanned and searchable microfilm ended, and 1995 when digital text archives began, those who wanted to find specific coverage of a person or event in the newspaper had little to go on. The Duluth Public Library indexed News Tribune stories for a range of years: 1893-1929, 1978-

1995. It is not comprehensive by any means. As the description indicates, *"Only articles of regional interest are included. The index does not include wire-service stories on national or international subjects, very short articles, or most sports stories."* The period of time between 1929 and 1978 is, as the Duluth Public Library reference librarian referred to it, "an indexing black hole." (Aho, 2015)

Interestingly, the 1893-1929 and 1988-1995 index cards were typed up and made into online searchable databases. Stories published between 1978-1988, however, will require going to the library to look at their bound, printed version of the index.

MICROFILM - Analog: Capturing the entire publication on microfilm is a suitable alternative to having the actual newsprint product, but it has several disadvantages to it. First, you must use the microfilm on site at the memory institution - it is rarely available to be "circulated." Second, if the reel you want is lost or misplaced, there is no backup. Third, without knowing a specific date and page for what you are seeking, the act of browsing through microfilm is mind and eye numbing. But for those whose research task is to assess the coverage of news for a particular time period or who are seeking news content from an era not available in any of the other forms, microfilm is the only choice.

The history of when and how the News Tribune started their microfilm archiving is lost but at some point early volumes of the paper were photographed. Depending on the year a researcher is looking for, the availability of the microfilmed record of the newspaper is questionable.

The first year of the Duluth Tribune's publication as a weekly (first edition on May 3, 1870) is available on microfilm at the Minnesota Historical Society. Other libraries and "memory institutions" have various start dates for their microfilm: Duluth Public Library (*from 1871*), the News Tribune newsroom (*from 1881*), the University of Minnesota – Duluth (*from 1890*) and the Wisconsin Historical Society's collection from October 1892 to April 1897.

MICROFILM - Scanned / Digital: Many newspapers have had their microfilmed publications scanned, enabling digital search and retrieval. For researchers who know a specific date and for those wanting to find stories on the topics or names for which they are searching, digital microfilm is a remarkable resource.

In the case of the Duluth News Tribune, the Readex / Newsbank partnership to acquire and scan newspaper microfilm has made available publications from May 16, 1881 to December 31, 1922 (which is the date before which newspapers are in the "public domain" under U.S. Copyright law).

The scanned microfilm does not just display the page, it has also been digitized so that the news content can be searched. But identifying the availability and location of these collections of digitized microfilm archives can be confusing.

The University of Minnesota Library has access to the Readex / Newsbank service called **America's Historical Newspapers.** The version they subscribe to includes **Early American Newspapers Series 1, 6 and 7, 1690-1922.** This provides access to hundreds of scanned historical newspapers. The Duluth News Tribune's scanned

microfilm, however, is part of **Early American Newspapers Series 2 and 3** and, therefore, is unavailable from the University's collection.

Once this type of material is scanned, the information companies will re-market and re-brand for different audiences. The Duluth News Tribune's content from 1876 to 1922 is part of Newsbank's *GeneologyBank*. As the librarian at the Duluth Public Library indicates, the cost of providing this kind of resource can be prohibitive.

> *Our library does not purchase this database, but many genealogy research sites do. Access to those sites is less expensive for individuals than for libraries, so our patrons often have better access than we do. We used to be able to use the GenealogyBank access as a sort of index; you could search but not view the content. That is no longer available.* (Aho, 2015)

DIGITAL ARCHIVES OF THE PRINT PUBLICATION: The Duluth News Tribune's digital archives of the print newspaper are available in-house (for the use of the newsroom) and commercially through NewsBank from January 11, 1995. The ease of searching these text-only archives is helpful for those who just want to find news stories on a particular name or topic, but having the text stripped from any sense of the placement and play of the story is not helpful for those who want to see the news in the context in which it was published. This would require a two-step process - locating the date and then going to microfilm.

Gale's Student Resources in Context has a sparse collection of News Tribune stories dating back to October 2000. But the lack of complete citation (there is no indication of section or page number where the story) would make that second step of location on microfilm all the more time-consuming.

WEB PUBLICATION: The early days of online publication for most news organizations is lost. Sometimes it was a change in publishing platform that wiped out an existing web archive. In other cases the news organization didn't consider the potential need for accessing past web content. The Duluth News Tribune added a change in ownership to the situation that has resulted in the loss of web content from its start in 1996 up until 2008.

For researchers who would like to see what was published on the website for a certain date, they might be lucky and find the Internet Archives' Wayback Machine captured the pages, but chances are poor. Not only is the capture sporadic (over the past 17 years of web publishing - 6200 days - there have been only 257 captures of the Duluth website), sometimes the link connects not to the historic capture but to the current website.

Locating stories published on the website relies on the webpage search engine which searches not for words in the story text, but words anywhere on the webpage. For the most part, search engines on websites lack sophisticated search prompts and can be frustrating for researchers. But if you find the story you want, it does display the website's placement and play of the content.

PDF PUBLICATION: Yet another version of the printed news product is available electronically with the News Tribune's PDF of the day's newspaper. It is searchable, similar to scanned microfilm, but, unfortunately, only the most current eight days of publication are available. The vendor, Olive Software Inc., offers its customers an "archive" feature for PDF versions but the News Tribune does not subscribe to that option. If it did, researchers would have as easily searched and fully representative

access to a newspaper publication in recent years as the NewsBank scanned microfilm provides for publications before 1923.

This complicated landscape of news access challenges even those charged with organizing and maintaining news collections. In the process of tracking down the various forms in which news publications and their contents can be accessed, the librarian for historic archives at the University of Minnesota-Duluth, referring just to an inventory of their own holdings, said, "For the record, compiling this information took me most of a morning. It certainly is chaos!!" (David, 2015)

Conclusion

Access to yesterday's news, whether it appeared in print or as ones and zeroes, will always be important for researchers and community members. But there is no guarantee that access will be available. Even the largest, most well-established vendors of news databases do not function as "archives." Witness the experience of the news librarian at the San Antonio (Texas) Express-News:

> *It was recently brought to our attention that our older stories (prior to 2006) were missing in the larger group files on Nexis. We then asked Nexis about it. The answer: About a year ago, Nexis decided to update those large group files to increase speed and relevancy. As a result, some publications had their stories older than 2006 removed from these files. They can still be found in our publication's file and in our statewide file. Of course, this does not take into account users who never drill down below All News or US Newspapers, but perhaps they don't need to search beyond 2006.* (Domel, 2014)

News organizations have always been preoccupied with creating the "daily miracle" that is the newspaper. Concern about what happens after the presses run or after the "publish" button is pushed in the CMS is less important. Creating and

maintaining a true archive, which requires a long-term preservation strategy and professional attention, takes money and time - two things in short supply in 21st century newspaper organizations. It is no wonder that the best descriptor for most news archives - print or digital - is "chaotic."

It is too late to recover the early history of news website publication. Both legacy and born-digital news operations have lost their opportunity to document their contributions to the new information environment in which we find ourselves. As early as 1997 at the 63rd IFLA Council and General Conference, Terry Kuny was warning of a "digital dark ages," an all-too prescient analysis of what was unfolding in the online news industry at that very point in time. Referring to the intellectual property and licensing regimes that restrict what can be held by libraries, Kuny, in the midst of his pessimistic presentation, optimistically hoped that newspapers' (and other types of publications') corporate owners might fill the preservation need. That optimism was misplaced (Kuny, 1997, p.3)

Eighteen years later, we may be at a moment in news archive preservation development where a critical mass of key players is forming. Certainly, memory institutions are aware of the issues surrounding digital preservation. Researchers in many disciplines are starting to recognize that a major source of primary data about a place and time in history is at risk. News database vendors and software companies may be eyeing future business opportunities. And finally, at long last, news organizations themselves may be awakening to the implications of the naive assumption that "everything on the Internet lives forever, right?" (Sillesen, 2014) Let's

hope the convergence of all of the key players leads to economic, technological,

organizational and societal solutions to ensure that we are not, indeed, living in a digital

dark ages.

Duluth News Tribune	1870s	1880s	1890s	1900s	1910s	1920s	1930s	1940s	1950s	1960s	1970s	1980s	1990s	2000s	2010s	Today
Pub/Hard Copy			hhh													
Story/Hard Copy					p p p	p p p	p p p	p p p	nnnnnnnp p p	nnnnnnnp p p	nnnnnnnp p p	nnnnnnnp p p	nnnn p p p	p p p	p p p	p
Pub/Microfilm	cccccccc	cccccccc	cccccccc	cccccccc	cccccccc	cccccccc	cccccccc	cccccccc	cccccccc	cccccccc	cccccccc	cccccccc	cccccccc	cccccccc	cccccccc	c
Pub/Digital MF		dndndndn	dndndndn	dndndndn	dndndndn	dn										
Story/Database													dndn	dndndndn	dndndndn	dn
Pub/Web-Internal													nn	nnnnnn		n
Pub/Web-External													w	w w w	w w w	w
Pub/PDF																c
Finding Aids				p p p p	p p p p p p	p p p p p p					p p	p p p p p	p p p			

h = Historical Society
n = Newsroom
p = public library
c = commercial product
dn = database - newsbank
dg = database - gale
w = web (ie: wayback)

ENDNOTES

Aho, K. (2015). Personal communication with co-author Paul.

Carlson, D. (1999-2009). *David Carlson's Virtual World: The Online Timeline*. Retrieved

from http://iml.jou.ufl.edu/carlson/1995s.shtml#1995

David, M. (2015). Personal communication with co-author Paul.

Domel, J. E. (2014) Posting to the NewsLib mailing list from September 10, 2014.

newslib@listserv.unc.edu

Gabriele, S. (2014). Transfiguring the Newspaper. *Amodern 2: Network Archaeology*.

Retrieved from htttp://amodern.net/article/transfiguring-the-newspaper.pdf

Gajda, T. (2015). Personal communication with co-author Paul.

Krueger, A. (2014) News Tribune Attic. Retrieved from http://attic.areavoices.com/

Kuny, T. (1997) A Digital Dark Ages? Challenges in the Preservation of Electronic

Information. *IFLA Conference Proceedings*, (1-12)

Larson, R. (2015). Personal communication with co-author Paul.

Metcalf, J. (1993). The Electronic News Library. In B. P. Semonche (Ed.), *News Media

Libraries: A Management Handbook* (220-231). Westport, CT: Greenwood Press.

Mills, T. F. (1981). Preserving Yesterday's News for Today's Historian: A Brief

History of Newspaper Preservation, Bibliography and Indexing. *Journal of Library*

History 16 (3), 463-487

Rothman, J. (1966). Preserving the News That's Fit to Print. *Indexer 5* (1),

   39-42.

Salmon, L. M. (1923). *The Newspaper and the Historian*. New York, NY: Oxford

   University Press.

Semonche, B. P. (1993a). News Library History. In B. P. Semonche (Ed.), *News Media*

   Libraries: A Management Handbook (1-45). Westport, CT: Greenwood Press.

Semonche, B. P. (1993b). Newspaper Indexing Policies and Procedures. In B. P.

   Semonche (Ed.), *News Media Libraries: A Management Handbook* (373-412).

   Westport, CT: Greenwood Press.

Sillesen, L. (2014) Minus proper archives, news outlets risk losing years of backstories

   forever. *Columbia Journalism Review*. Retrieved from

   http://www.cjr.org/behind_the_news/minus_proper_archives_many_new.php

Skinner, K. and Schultz, M. (2014). *Guidelines for Digital Newspaper Preservation*

   Readiness. Atlanta, GA: Educopia Institute.