

## MULTIMEDIA REVIEW

### A Survey of Online Digital Newspaper and Genealogy Archives: Resources, Cost, and Access<sup>1</sup>

Digitized historical newspapers opened the world of professional women whistlers to me. What began as a curiosity became a preoccupation as simultaneous keyword searches of hundreds of newspapers revealed details and a broad context for the history of these forgotten musicians. These women inspired the publication of hundreds of items including reviews of their performances, articles on the propriety of their activities, portraits, advertisements, announcements of coming events, short stories, and poetry and rhymes immortalizing the whistling girl. Although any one piece of information might have little value on its own, the volume of material begins to form a historical narrative, one that emerges with a speed that would not have been possible in the analog age.

The process of searching shaped my understanding of digitized historical newspapers as a genre of content and the Internet as a point of access and delivery. As someone new to using digitized historical newspapers, I first had to recognize the breadth of the resources and identify the databases that held documents of potential value. No single database, free or otherwise, provides comprehensive access, although some come closer than others. Moreover, a number of valuable historical newspapers are preserved in the relatively small collections of cultural heritage organizations whose online resources are isolated and difficult to search. To explore the richness of these resources, some larger sense of the potential for research is necessary. This review essay charts this landscape and suggests avenues to access.

Commercial databases provide rich holdings of digitized historical newspapers; however, despite the important service they provide, their expense raises concerns about access to documents preserving cultural heritage.<sup>2</sup> America's Historical Newspapers, ProQuest Historical Newspapers, GenealogyBank.com, Newspaper-ARCHIVE.com, Newspapers.com, and Paperofrecord.ca preserve huge amounts of historical material, but only for those with appropriate institutional affiliations, proximity to a subscribing institution, or a willingness to pay for access as an individual. Although some databases hold materials in common, each offers something unique; thus, to choose one over another is to eliminate possibilities for discovery. And yet financial realities, and the scarcity of research institutions that provide

<sup>1</sup> The resources surveyed here are: America's Historical Newspapers, [www.readex.com/content/americas-historical-newspapers](http://www.readex.com/content/americas-historical-newspapers); Chronicling America, <http://chroniclingamerica.loc.gov/>; Digital Public Library of America (DPLA); <http://dp.la/>; Genealogy Bank, [genealogybank.com](http://www.genealogybank.com); Google News, [news.google.com](http://news.google.com); NewspaperARCHIVE, [newspaperarchive.com](http://www.newspaperarchive.com); Newspapers.com; Paperofrecord.ca; ProQuest Historical Newspapers, <http://www.proquest.com/en-US/catalogs/databases/detail/pq-hist-news.shtml>; "Wikipedia: List of Online Newspaper Archives," [http://en.wikipedia.org/wiki/Wikipedia:List\\_of\\_online\\_newspaper\\_archives](http://en.wikipedia.org/wiki/Wikipedia:List_of_online_newspaper_archives).

<sup>2</sup> These concerns are documented in "Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities," *Open Access, Max-Planck Gesellschaft*, 22 October 2003, <http://oa.mpg.de/lang/en-uk/berlin-prozess/berliner-erklarung/>.

comprehensive access to commercial sites, make choices necessary, increasing the need for a broad sense of the scope and limitations of these databases.

America's Historical Newspapers is one of the richest commercial sources of digitized historical newspapers. It began in the mid-twentieth century as a collaboration between Readex Microprint Corporation of Chester, Vermont and the American Antiquarian Society in Worcester, Massachusetts, with the mission to preserve and distribute historical newspapers as microprint. Now digitized, the assets of this collection have expanded to include materials from over ninety institutions, with new acquisitions managed by a board of historians and bibliographers with academic appointments.<sup>3</sup> Institutions can license discrete units or entire collections, and can limit their access by series, state, or a selection of newspapers in order to serve a specialized audience of library patrons or conform to a limited budget. Series available within America's Historical Newspapers include Early American Newspapers 1690–1922; 20th-Century American Newspapers (since 1923); American Ethnic Newspapers (including about 270 African American newspapers, 1827–1998; and hundreds of Hispanic-American newspapers, 1808–1980); and American Newspaper Archives, a regionally diverse group of ten long-running urban newspapers, including the *Augusta Chronicle*, *Baton Rouge Advocate*, *Cleveland Plain Dealer*, and *Dallas Morning News*. With access to the entire database, it is possible to search publications from all fifty states and the District of Columbia from 1690 nearly to the present, but such access is restricted to institutions whose mission and resources allow for it.<sup>4</sup>

The content of America's Historical Newspapers is available to individuals as monthly or yearly subscriptions through GenealogyBank.com, a database marketed to genealogists by NewsBank, the proprietor of Readex since 1984. GenealogyBank.com includes over 6,000 newspapers published in the United States between 1690 and 2010 from all fifty states and the District of Columbia, and, like America's Historical Newspapers, the number of publications available continues to grow.<sup>5</sup>

Readex has also collaborated with the Center for Research Libraries (CRL) to create the World Newspaper Archive, guided by many of the same structures and principles employed by America's Historical Newspapers: a board of historians and information specialists, and a set of collaborating institutions that provide documents to digitize and resources to support the process. They currently offer African Newspapers, 1800–1922; Eastern European Newspapers 1835–1922; Latin American Newspapers, Series I and II, 1805–1922; and South Asian Newspapers, 1864–1922. The impetus for creating such collections is preservation as well as insuring “persistent and affordable access” for members of the Center for Research Libraries who subscribe to these collections.<sup>6</sup>

<sup>3</sup> See the history of Readex at <http://www.readex.com/page/history-readex>.

<sup>4</sup> Readex supplies a list of these available resources at <http://www.readex.com/content/americas-historical-newspapers>.

<sup>5</sup> GenealogyBank's historical newspapers can be accessed at <http://www.genealogybank.com/gbnk/newspapers/>. This page also has a link to a list of the newspapers it holds organized by state of publication.

<sup>6</sup> For more information on the World Newspaper Archive, see <http://www.crl.edu/collaborative-digitization/world-newspaper-archive>.

ProQuest Historical Newspapers holds yet another set of more than forty digitized historical newspapers and over thirty million digitized pages. Its content includes mainstays such as the *New York Times*, *Detroit Free Press*, and the *Los Angeles Times* as well as international, African American, and American Jewish newspaper titles. Like America's Historical Newspapers, ProQuest markets discrete portions as well as the entirety of its collection to research libraries.<sup>7</sup> ProQuest also provides individuals access to its historical newspapers for a price. The researcher can visit each newspaper website and search historical newspapers through the link to the archive, which is usually found on the site map. Once citations are found, a credit card provides immediate online access, with the cost varying dependent on the particular newspaper, the period of access licensed, and the number of individual items selected to be downloaded.<sup>8</sup> Once the fee is paid, the articles are available as easily downloadable PDFs.<sup>9</sup>

NewspaperARCHIVE.com, recently acquired by ProQuest, provides yet another point of access to digitized historical newspapers through institutional or individual subscription.<sup>10</sup> It boasts over 120 million pages of newspapers, including over 5,000 titles spanning 1607 to present, from all fifty states, the District of Columbia, and a host of countries outside the United States, with strong holdings from Canada, Jamaica, and the United Kingdom.<sup>11</sup> It offers a free trial period so that the researcher can get a sense of whether its offerings are worth a subscription, and allows for an initial search to introduce potential users to the resource. In my case, it found 402,325 hits across three centuries on a search of "whistling"; the phrase "whistling girls" brought 120,389 hits, but only allowed me to see the results with a trial membership requiring my credit card information. It also offers a set of free newspapers to search. Categories such as "Arts and Entertainment," "Natural Disasters," and "Surnames" invite exploration. Having been marked by Google's automated "spiders," which "crawl" content on the "web" to make it searchable, a Google search will provide hits from the database, but after the initial article is found a deeper investigation requires a subscription or trial membership.

Newspapers.com is the newest of these commercial databases, launched by the publicly held Ancestry.com in November 2012.<sup>12</sup> As with other databases marketed

<sup>7</sup> For a list of the historical newspapers made available by ProQuest, see <http://www.proquest.com/en-US/catalogs/databases/detail/pq-hist-news.shtml>.

<sup>8</sup> For instance, the *Detroit Free Press* provides a "10 Article 24-hour Pass" for \$11.95; a "one month 25 Article Pack" for \$59.95; a "one month 10 Article Pack" for \$24.95; a "one week 3 Article Pack" for \$9.95; or a single document purchase for \$3.95. The *Chicago Tribune* offers similar options, including an annual pass with access to 360 articles, a "power user" annual pass with access to 1,200 articles, as well as monthly, power-user monthly, weekly, twenty-four-hour, and single-article options.

<sup>9</sup> James Mussell has contributed an insightful review of ProQuest Historical Newspapers to *Reviews in History: Covering Books and Digital Resources across All Fields of History*, 23 June 2011, <http://www.history.ac.uk/reviews/review/1096>.

<sup>10</sup> See ProQuest's press release announcing the acquisition at <http://www.proquest.com/en-US/aboutus/pressroom/13/20130311.shtml>.

<sup>11</sup> See <http://newspaperarchive.com/>.

<sup>12</sup> The press release announcing the new database is fascinating in part for its explanation of "forward-looking statements," such as the projected benefits to its subscribers and plans for growth of the database's content. See <http://corporate.ancestry.com/press/press-releases/2012/11/ancestry.com-launches-new-web-site-newspapers.com/>.

to genealogists, it offers a weeklong trial for the curious with a credit card—the card is charged at the end of the week should the membership not be cancelled. The search and presentation tools operate smoothly, with all the advantages of the most recent search and presentation technology. Although its holdings are immense, with over 2,000 newspaper publications and fifty-seven million pages to search, it is still small in comparison to NewspaperARCHIVE.com and GenealogyBank.com. Although I had hoped that its resources would overlap with those of the other databases, making a subscription unnecessary for an attempt at comprehensiveness, a quick search of “whistling girls” had over 800 hits, a number of which were from newspapers I had not yet encountered.

The history of Paperofrecord.ca, which claims to be the first Internet site for digitized newspapers, dramatizes the potential and the risks associated with the commercial collection and distribution of digitized historical newspapers. It began with the mission of digitizing newspapers from around the globe and making them accessible online to the general public, starting with the *Toronto Star* in 1999. The database later became known among historians as a site to access large collections of digitized historical newspapers from Mexico and Canada.<sup>13</sup> In late 2008, however, Google acquired the site, and the materials disappeared as Google added its contents to Google News Archive, inconveniencing numerous users dependent on the resources.<sup>14</sup> The website reappeared in 2011 after Google abandoned Google News Archive, and now provides access to its original holdings. The delivery system reveals its age, in that each newspaper must be searched independently, and the search term is not always highlighted on the pages found. In addition, the content is surprisingly small considering its global ambitions and compared with commercial sites.

Google News Archive still offers free access to the set of newspapers that it collected and crawled while it was building its archival collection of historical newspapers.<sup>15</sup> Although the homepage has been eliminated, all the material remains, and can be searched through Google News.<sup>16</sup> Google’s collaborative efforts have resulted in a hybrid of free and pay-per-view newspaper access. In a sense, the Google newspaper site now serves as an informal search engine for the ProQuest Historical Newspapers, though it can provide only a citation and a link to an abstract for free.

An abundance of sites offering free access to digitized historical newspapers has developed parallel to pay services. *Chronicling America* represents the standard to which freely accessible projects now aspire.<sup>17</sup> The website’s development began in

<sup>13</sup> For the story of its origins, see the Paper of Record website <https://paperofrecord.hypernet.ca/default.asp>. For a list of its holdings, see <https://paperofrecord.hypernet.ca/papers.asp>.

<sup>14</sup> See the discussion of this lack of access to newspapers once held by Paper of Record and responses from Google at <http://productforums.google.com/forum/#!topic/news/2Bdyv-lk1RE>.

<sup>15</sup> Sources conflict as to whether the archive began in 2006 or 2008. A blog about the site from 2010 can be found at <http://googlesystem.blogspot.com/2010/03/browse-newspapers-in-google-news.html>.

<sup>16</sup> A list of available newspapers alphabetized by title can be found at <http://news.google.com/newspapers>.

<sup>17</sup> See, for instance, Katherine Skinner and Emily Gore, “The MetaArchive Cooperative: Chronicles in Distributed Digital Preservation,” in “2010 Conference Proceedings from US Workshop on Roadmap for Digital Preservation Interoperability Framework, ACM Digital Library,” October 2011.

2004 as the National Digital Newspaper Program (NDNP), a collaboration between the Library of Congress (LC) and the National Endowment for the Humanities (NEH) that supports further growth through annual grants. In the beginning, six geographically diverse institutions established a foundation of images for the project; it opened to a general audience in 2007. In its first stage, the images represented newspapers published between 1900 and 1910, mostly from California, Florida, Kentucky, New York, Utah, and Virginia, and Washington D.C., where the six institutions and the Library of Congress are located.<sup>18</sup> Coverage now spans from 1836 to 1922 and includes 6,673,511 images from 1,105 newspapers published in thirty-five states and the District of Columbia.<sup>19</sup> Like many other digital historical archives, the NDNP project uses 1922 as a terminus date for the collection, which assured in the early days that the material fell into public domain. Despite the passage of time, there are currently no plans to expand the collection criteria beyond 1922, an example of how concerns about issues of copyright limit the content of such sites.

When awarding grants, the NDNP privileges applicants who have publicly available master microfilms in complete runs, or with runs that can be made complete with images from the original paper sources. They seek to preserve orphaned titles—newspapers whose copyright holders cannot be identified or contacted—of high research value, publications that represent the activities of significant minority communities, and “papers of record,” those that distribute legal notices, news of state and regional government affairs, and announcements of community news and events, including births, deaths, and marriages.<sup>20</sup> Although the collection continues to grow, it already represents a rich source of easily searched materials, beyond the dreams of anyone who in past decades depended on microfilm for remote access.

In addition to providing images of historical newspapers with metadata (descriptive information about the original newspaper and its surrogate, the digital image), the website includes a database of bibliographic records for 151,797 newspapers published in the United States from 1690 to present and available on microfilm, a resource established through the United States Newspaper Program (USNP), the predecessor to the NDNP that spearheaded the preservation of U.S. newspapers.<sup>21</sup>

A number of freely accessible, collaboratively constructed sites support the search for digitized historical newspapers by country, by state, or across cultural heritage organizations. Wikipedia offers one such useful collection of links in its entry, “List of Online Newspaper Archives.”<sup>22</sup> The materials are arranged alphabetically by

---

Skinner and Gore assert that *Chronicling America* represents a “replicable archival model for digital newspaper preservation” that addresses the needs of cultural memory organizations to preserve their digital historical newspaper collections.

<sup>18</sup> Information on NDNP is available at <http://www.loc.gov/ndnp/>.

<sup>19</sup> For information on applying for a grant from the NDNP, see <http://www.neh.gov/grants/preservation/national-digital-newspaper-program>.

<sup>20</sup> For a description of the NDNP’s selection criteria, see <http://www.loc.gov/ndnp/guidelines/selection.html>.

<sup>21</sup> The content of *Chronicling America* is listed at <http://chroniclingamerica.loc.gov/search/titles/>.

<sup>22</sup> See Wikipedia: List of online newspaper archives, [http://en.wikipedia.org/wiki/Wikipedia:List\\_of\\_online\\_newspaper\\_archives](http://en.wikipedia.org/wiki/Wikipedia:List_of_online_newspaper_archives).

country; the United States is broken into subcategories of individual states. Each link to a newspaper or site indicates if access is free or for a fee. Included are collections of the highest quality and greatest ease of searching as well as those that are difficult to use and that are not well documented. As with many other user-constructed resources online, this list is subject to change and may disappear. Because it is not Wikipedia's mission to collect such links, which support access to other resources rather than providing encyclopedic information, its appropriateness was discussed in 2010 on its "talk" page, where contributors suggest changes to an entry's content, and its removal considered.<sup>23</sup> The list's existence, however informal and precarious, provides one solution to the isolation from the mainstream of digital resources maintained by small cultural heritage institutions and individuals. A need exists for a site that provides access in one easily searchable location.

In the long term, the Digital Public Library of America (DPLA) provides a potential solution to the problem of isolated resources. Released in a beta version in April 2013, the DPLA ([dp.la](http://dp.la)) represents one of the newest efforts to maintain free access to heritage materials, to which it serves as a portal. To date, over 5.5 million records at DPLA link to digital objects on the content providers' sites. In a much more formal process and on a grander scale, the DPLA does the work of the Wikipedia article on digitized historical newspaper archives, but also seeks to include "the full breadth of human expression—written word, works of art and culture, records of American heritage, efforts and data of science."<sup>24</sup>

Finding newspapers through DPLA can be complicated. The search engine is not yet so sophisticated as to allow users to limit searches to newspapers. But searches inevitably lead to other sites that have greater specificity of search capacity. Because of the variety of cultural objects DPLA makes available, a search for "newspaper" results in records of artistic photographs of newspapers, books and articles about newspapers, clippings in archival collections, and other documents related to the subject "newspaper." General searches often require repeated refinement, but are worth the time if the papers of interest are only found in small public libraries. When these institutions contribute their metadata to DPLA, they become more visible to the general public, even if the process of finding the resources through DPLA is still somewhat serendipitous.

The potential of DPLA as a collection of metadata can be seen through a search for "digitized historical newspapers." This phrase generates 3,359 hits, which may be refined by format (text or image, in this case), contributing institution (holder of the digitized item), partner (state institution collaborating with DPLA), location (cities, counties, and states in the United States), and subject (e.g., United States, Texas, Places, Newspapers, Journalism). This search led to "The Portal to Texas History" (<http://texashistory.unt.edu>). This new site allows users to limit

<sup>23</sup> The second tab from the left at the top of every Wikipedia article is marked "talk." When readers clicks the tab with a mouse, they can read these discussions, which often inform the shape that an article takes.

<sup>24</sup> More information on the Digitized Public Library of America is available at its website: [dp.la](http://dp.la). The Berkman Center for Internet & Society at Harvard University recently posted an article describing its role in the creation of the site. See <http://cyber.law.harvard.edu/research/dpla>.



searches to newspapers. A search of “whistling girls” generated fourteen hits, one of which related to an article in the *Waco Evening News* from 1889 with new information. (Never whistle in a dressing room, or you will whistle the actress in the neighboring room out of the theater. Sing instead.) A search using a slightly different phrase, “historical newspapers,” leads to the Georgia Newspaper Project, which offers links to nine different archives of the state’s newspapers produced (<http://dlg.galileo.usg.edu/Institutions/gnp.html>). Depending on the keywords, a single search on DPLA may identify a number of partners or contributing institutions that hold digitized newspapers, such as Georgia Historic Papers, which holds the *Cherokee Phoenix*, the *Dublin Post*, and the *Colored Tribune*; Cabarrus County Public Library in North Carolina made available a run of the *Concord Standard* from 1888 to 1899 with the support of the North Carolina Digital Heritage Center. The resulting website is called DigitalNC (<http://www.digitalnc.org/>), and represents one of many websites and data hubs that result from partnerships with DPLA.

Digitized historical newspapers are an important element of cultural heritage around the world. These newspapers are preserved not only for scholars, but also for members of the general public who have an interest in understanding their family’s origins and circumstances as well as accessing information about the past more generally. Although researchers of all sorts still use microfilms of historical newspapers, the slow and cumbersome process is becoming rarer as full runs of thousands of publications have been digitized and made available on the Internet. Although the process has made historical newspapers an increasingly valuable and accessible, it has at the same time raised new challenges. Many of the sites are dynamic, with new material added regularly; commercial sites change hands, search engines are upgraded, new sites emerge and old sites disappear. As with any historical work, research is never complete, and there is always the sense that with access to just one more source the picture of the past might become clearer, if not more complicated.

Maribeth Clark