Digitizing Newspaper Collections: A Shift from Microfilm to Digital, National & State Newspaper Digitization Programs

Rachael Altman
IST 677
March 26, 2011
Digitizing Newspaper Collections: A Shift from Microfilm to Digital, National & State Newspaper Digitization Programs

Introduction

The digitization of newspapers is a fairly new form of access and preservation. Persons and institutions involved in the digitization of newspapers are divided when it comes to deciding to preserve newspapers through microfilm or in a digital format. There are several nationwide newspaper digitization efforts that have created and established standards and methods for the digitization process. Despite the fact that the digitization of newspapers is a recent development, there is a vast amount of literature dedicated to the topic. Newspaper digitization programs and projects occur at libraries and research institutions, as well as on the state level, national level, and international level. Perhaps the main purpose of newspaper digitization is to provide the public with free access to historical documents. Aside from free, open access, the purpose of newspaper digitization is to preserve the historic documents into the future. This literature review discusses the recent shift from microfilm to digital forms, the advantages and disadvantages of digitization, the standards and methods of digitization, and digitization programs on the state and national level.
Microfilm vs. Digital

The “Guidelines for Digitization” notes that microfilm is the most accepted standard for preserving newspapers. Creating the microfilm of the newspaper is the first step in the digitization process. Further, it should be recognized that filming and digitization can happen simultaneously because scanning from microfilm is faster and less expensive, and newspaper digitization is mostly done from microfilm. However, print can be used and microfilm quality may present challenges due to older film, blurry film, or incomplete film (Library of Michigan, 2011, p. 1).

According to Hakli, “In terms of preservation, digitization can be defined ‘as a method to improve access to otherwise fragile or rare analogue materials. In contrast we could transfer digital materials such as simple text and pictures to paper or microfilm formats, which are safer than any file format today’” (quoted in Hasenay & Krtalic, 2010, p.248).

However, Hedstrom defines digital preservation as “the planning resource allocation, and application of preservation methods and technologies necessary to ensure that digital information of continuing value remains accessible and usable” (2010, p. 248). Furthermore, Keith Rajecki, author of “Preservation and Archiving Solutions” notes “the digitization and archival preservation of important historic artifacts is often made available through the internet, enabling scholarly research, exhibitions, and other research and educational activities” (2010, p.33).

The article, “Preservation of newspapers: Theoretical approaches and practical achievements”, by Damir Hasenay and Maja Krtalic, discusses the two methods of preservation: microfilming and digitization. Microfilming was the original method for preserving newspapers, but microfilming and digitization are beginning to interweave, and digitization is starting to be considered more often as a means of preservation for newspapers. The methods of digitization and microfilming for newspaper preservation can be combined because newspaper digitization is most often continued from newspaper microfilming. Additionally, newspaper content can be reformatted through digitization (Hasenay & Krtalic, 2010, p. 248).

Similarly, Ray L. Murray, author of the article “Toward a Metadata Standard for Digitized Historical Newspapers”, notes that digitizing from microfilm is an efficient way to capture a high volume of data. Although in the end what is created is a digital image of the original page, the characteristics of the intermediary medium of the film should not be ignored. The content will have been transferred three times: once to the film, once to the print negative, and once when being digitized (2001, n.p.).

Additionally, according to the University of Illinois Library, converting microfilm to digital formats may be impossible if the original film is in poor condition, if the original film cannot be read by OCR, or if the film is unfocused or varied in density (University of Illinois Library, 2010).
Advantages & Disadvantages of Digitization

The advantages to digitization seem clear: better use of content, global access for all users with Internet capability, greater access for research, more innovative approaches to indexing, standardized metadata and cataloging, less physical handling of the original print copy or original microfilm, and greater usage and copying without degradation.

Advantages may also include great opportunities for technological growth, collaboration between institutions, and cultural heritage preservation. According to the Wyoming State Digitization Plan, the state of Wyoming decided to create a digital history collection in order to 1) increase awareness of Wyoming’s histories and how they connect to the nation and the world; 2) provide alternative access to primary sources for education and research purposes through digital technologies; and 3) increase cultural tourism for economic development in the region (2005, p. 1).

While there are numerous advantages to digitization, there are even more disadvantages according to Damir Hasenay and Maja Krtalic, authors of “Preservation of newspapers: Theoretical approaches and practical achievements”. It should be noted that the preservation of digital materials demands far more active management, especially considering the frequency of changes and the fact that permanence may become doubtful in the digital realm.

Further, there are greater amounts of technical and organizational obstacles. Technical issues include, instability of media on which digital information is stored and the instability of the technology needed for its use, instability of location, and deterioration of digital media caused by damage to the materials that make the data layer. Technical obsolescence caused by development and changes in the digital world is also a challenge because as technology develops, machines and the way in which information is processed, stored, presented and read also changes. Organizational challenges include, management of materials, human and financial resources, legal concerns, society awareness, and cooperation on national and international levels. Disputes of clearly defined goals, methods, techniques, standards, and guidelines are still developing in the digitization world. The legality of digitization in terms of copyright is an issue because newspaper articles may have multiple copyright holders—the reporter, the photographer, and the publisher (Hasenay & Krtalic, 2010, p. 249).

Additionally, the International Coalition on Newspapers (ICON) recognizes the challenges of the digitization of newspapers. Some of the challenges include the fact that companies such as The New York Times distribute “backfiles” directly through the Internet, which leads to libraries no longer being able to rely on micropublishers such as ProQuest and LexisNexis. This also leads to serious decision making about digitization and maintenance of print and microfilm collections (2010, n.p.).

The United States Newspaper Program serves as a pioneer to newspaper digitization because other countries are only beginning to make the shift from microfilm to digitization. Patricia Liebetrut, author of “Newspaper Digitisation in South Africa” discusses the positive and negative aspects of microfilm and digitization: “The conclusion is that scanning from microfilm is and can only be as good as the original
microfilm. Unless international standards have been applied (and these have changed over the years) in microfilm capture, the condition may not be the best. However it is sometimes the only resource available for scanning old and fragile newspapers” (Liebetrau, 2010, p. 18).

Similarly, Deborah Novotny, author of “Newspaper Digitisation: A Silver Bullet”, discusses the British Library’s shift from microfilm to digital. The shift from microfilm to digitization is more enticing because digital copies have color, rather than the old fashioned black and white microfilm. Further, microfilm is more difficult to access, while digital copies are easily accessible from a multitude of locations. Novotny notes that the main benefit of digitization is immediate access from anywhere, by anyone, free of charge. However, there are also risks, including copyright, digital storage maintenance, proven metadata, the production of digital surrogates, inclusion in national and international registries, and the cost benefit of the digital life cycle (2010, p. 25).

**Standardization & Methods**

The Cornell University Library Research Group notes that quality control (QC) “encompasses procedures and techniques to verify the quality, accuracy, and consistency of digital products and is essential to all phases of a project life cycle to ensure that the product meets preset standards and goals” (quoted in University of Maryland Library, 2007, p. 13).

According to “4.0 Best Practices for Newspaper Digitization” developed by the University of Illinois Library, “the single most critical factor in the success of newspaper digitization is the availability of good quality microfilm. Although it is possible to digitize newspapers from an original print copy, this process is very labor-intensive and considerably more expensive than digitizing from film” (University of Illinois Library, 2010, n.p.).

Hasenay and Krtalic argue that the standards for the digitization of newspapers are still developing. Standardization is a systematic approach, which makes it difficult to find a balance between accessibility and feasibility within the limitations of available resources (2010, p. 250). Standards may also include paying more attention to one part of the collection over another. In the case of the Pennsylvania Digital Library, greater priority was given to the digitization of printed newspapers because such items represented local history and could be used for local history research (2008, p. 5).

Further, the “4.0 Best Practices for Newspaper Digitization” discusses the necessity to evaluate the quality and condition of the microfilm if using it for digitization. Film used for newspaper digitization should be a clean second-generation duplicate to silver negative film because it has better contrast and it is easier to correct scratches. The film should also be polyester rather than acetate because polyester is stable and durable. Generally speaking, microfilms produced pre-1980 are acetate stock; this can be identified if the film is brittle, blistered, or stinky (University of Illinois Library, 2010, n.p.).
Another important consideration for digitization is whether to digitize in-house or to outsource the project. According to the “University of Maryland Best Practice Guidelines for Digital Collections”, the decision to digitize in-house or via outsourcing depends on the “scope, nature, fragility, and uniqueness of the materials.” Further, the project budget is one of the most important documents for any digitization project or program (University of Maryland, 2007, p. 13).

The “Guidelines for Digitization” discusses why libraries and institutions should digitize paper or microfilm collections. Since libraries hold few copies of newspapers, and often the access is limited, digitization can globally increase access provided by the library or institution, and decrease the need for use of the original. Similarly, Skinner and Gore created a presentation for the MetaArchive Cooperative, which discusses the National Digital Newspaper Program’s focus on the standardization of newspaper digitization and establishing the best practices.

Prior to NDNP various digitization practices were used by individual institutions, yet NDNP and MetaArchive now study the needs of digital newspaper collections, the creation of preservation tools, the creation of “real” collections, the range of digital formats, the various systems for storage, metadata creation, and Optical Character Recognition (OCR) (Skinner & Gore, 2010). In relation to OCR and metadata, searchable and accessible digital images processed with OCR software make newspaper text easily accessible to users; the metadata is then created for each title, issue, image, and article based on the OCR scan.

Further, the article “Toward a Metadata Standard for Digitized Historical Newspapers”, by Ray L. Murray, discusses the need for structural metadata. In order to sustain and provide access in a national system, standardized metadata design is required. The creation of the NDNP lead to the creation of a universally accepted metadata standard, which is currently used by the NDNP and all of its Statewide Newspaper Digitization Projects.

The NDNP created the Metadata Encoding and Transmission Standard (METS), which is an “XML document format designed to handle complex objects, and to facilitate management of objects within a repository, or between repositories. The development teams designs separate METS document template titles, issues, and reels. Digitized titles have additional data—essays and geographic coverage data—included in the title record. Additionally, since newspapers come in different forms—microfilmed pages and the original print pages—each page will have multiple surrogates: the scanned TIFF file, a service image JPEG2000 or PDF, and the OCR text for the page” (NDNP, 2010, n.p.).

Overview of National Newspaper Digitization Programs

The National Digital Newspaper Program (NDNP) is a partnership between the National Endowment for the Humanities (NEH), the Library of Congress (LC), and state newspapers digitization projects to provide access to United States newspapers published between 1836 and 1922. NEH provides awards for state projects to select and digitize historically significant newspapers that are permanently maintained by the Library of Congress (NEH, 2010, n.p.).
As part of the National Digital Newspaper Program, Chronicling America: Historic American Newspapers is a prototype of the digital resources produced through NDNP. It is a free service provided to Internet users around the world. Users are able to search the digitized newspapers, as well as consult a national newspaper directory of bibliographic information and holdings information used to identify newspapers in various formats. Additionally, the NEH aims to represent every state and territory within the United States in the Chronicling America Program.

In relation to Chronically America, the Library of Congress Historic Newspapers Flickr Pilot Project is part of the National Digital Newspaper Program and the Chronicling America Collection; it allows users to explore “history’s first draft” by providing access to illustrated pages in old newspaper selection from the LC digital collections. Users are able to tag photos (creating user generated metadata) and make comments on the photos in the Flickr account. The images in the Flickr account come from the New York Tribune from the years 1909-1910 (Library of Congress, 2010, n.p.).

Further, the United States Newspaper Program (USNP) is a cooperative national effort among the states and the federal government to locate, catalog, and preserve newspapers published in the United States from the eighteenth century to the present. Funding is provided through the National Endowment for the Humanities, and technical assistance is provided through the Library of Congress.

The USNP has supported projects in all fifty states, the District of Columbia, Puerto Rico, and the U.S. Virgin Islands. Each project is orchestrated by an individual organization within a state or territory, which is usually the state’s largest newspaper repository. The items for the projects come from the holdings in public libraries, county courthouses, newspaper offices, historical museums, college and university libraries, archives, and historical societies. Catalog records are then entered into a national database, which is maintained by the Online Computer Library Center (OCLC), and access is provided to users through more than 53,500 computer terminals worldwide. Further, microfilm copies of the newspapers are available to researchers anywhere in the country through inter-library loan (USNP, 2010, n.p.)

**State Newspaper Digitization Projects Supported by the USNP**

While there are fifty-three State Newspaper Projects, this literature review will only highlight three unique projects: Alaska, Illinois, and Vermont. Each State Newspaper Project provides detailed documentation of the project phases—which generally include a Planning Phase, Implementation Phase, Microfilming Phase, and Digitization Phase—as well as how the items were catalogued, and how the items can be accessed.

**Alaska**

Between 1991 and 1998, the Alaska Newspaper Project (ANP) was a participant in the USNP. The ANP was primarily supported by grants from the National Endowment for the Humanities to the Alaska State Library, Division of Libraries, Archives, and Museums, Alaska Department of Education in Juneau, Alaska. Additional funding was
provided from the Rasmuson Library of the University of Alaska Fairbanks and the Consortium Library of the University of Alaska Anchorage. Between 1996-1998, the project in Alaska “microfilmed 100,000 pages of such papers as the Forty-Ninth Star and the Esquimaux, an 1866 newspaper for men laying Western Union's overland line. Included are handwritten newspapers from before the U.S. purchase of Alaska. The project has received $650,190 in NEH support” (NEH, 2010, n.p.).

Despite the fact that the Alaska Newspaper Project has microfilmed numerous pages of state newspaper history, the ANP has yet to make the shift to creating a completely digitized newspaper collection. The files have been catalogued in the Alaska Newspaper database, and the records newspaper items are available through the Alaska State Library. There are 209,002 digitized newspaper files available in the database. The files are searchable by the title of the publication, year, headline, and author (Alaska State Historical Library, 2008, n.p.).

Illinois

“The Illinois Newspaper Project has cataloged more than 20,500 titles, among them the Herald of the Prairies, the Griggsville Reflector, and the Ruralist. The project has received $5,176,621 in NEH support” (NEH, 2010, n.p.). The Illinois Newspaper Project is unique because it consists of three major digitization projects at the University of Illinois at Urbana-Champaign, the Chicago History Museum Research Center, and the Abraham Lincoln Presidential Library. Another unique aspect of the Illinois Newspaper Collection is that it provides collaborative access to digital newspaper collections statewide, including digital collections held at The University of Chicago, Bradley University, Illinois State University, Chicago Public Library, Metropolis Public Library, and Southern Illinois University.

In June 2009, the Illinois Newspaper Project at the University of Illinois at Urbana-Champaign received funding from NEH and NDNP to digitize 100,000 pages of historically significant Illinois newspapers dating from 1865-1922. The digitized newspapers will be contributed to the Chronicling America Program. This Project will digitize the following Illinois newspaper titles in entirety: Chicago Eagle (1982-1920), The Broad Ax (1895-1922), The Day Book (1911-1917), The Cairo Evening Bulletin (1868-1870), The Cairo Daily Bulletin (1870-1872), The Cairo Bulletin (1872-1878), The Daily Cairo Bulletin (1878-1884), and the Cairo Bulletin (1904-1922) (University of Illinois, 2009, n.p.).

The University of Illinois Library works directly with the Abraham Lincoln Presidential Library Newspaper Microfilm Collection, which is the largest microfilm repository in the state of Illinois. While the Abraham Lincoln Library has not digitized its collection, the University of Illinois Library created a database which houses the cataloged records of the entire newspaper microfilm collection at the Abraham Lincoln Library (2009, n.p.)

The Illinois Newspaper Project located at the Chicago History Museum Research Center is home to the Chicago Daily News Database, which includes photographs in the
Chicago Daily News Collection (1902-1933). Despite the collection of digitized photographs available online through Chicago History Museum, digitized print newspapers are not yet available. Users have access to view the cataloged records of the entire collection via the Internet, but users must pay a visit to the Chicago History Museum in order to physically access the newspapers. The fact that the Chicago History Museum and the Abraham Lincoln Library have yet to make the shift from microfilm to digital is proof that the digitization of newspapers is still a new technology not yet completely standardized or embraced (2011, n.p.)

Vermont

During 1997-2001, “the Vermont Newspaper Project (VTNP) produced 267,000 pages of microfilmed holdings including Vermont's first newspaper, the Vermont Gazette, and Green Mountain Post Boy, which was printed in 1781 by Judah Paddock Spooner and Timothy Green in Westminster on the historic Dresden Press (the first official printing press of the state of Vermont). Newspapers for Italian, French-Canadian, Slavic, and other immigrant groups working in the state’s quarries, lumberyards, and textile factories are also represented. The project has received $427,894 in NEH support” (NEH, 2010, n.p.).

As of July 14, 2010, the University of Vermont Libraries has received funding from the National Endowment for the Humanities in the amount of $391,552 to further support the Vermont Digital Newspaper Project as part of the NDNP. The University of Vermont Libraries will work collaboratively with the Vermont Department of Libraries, the Isley Public Library of Middlebury and the Vermont Historical Society to select, digitize, and provide access to up to 10,000 pages of Vermont newspapers published between 1836 and 1922. The digitized newspapers will be accessible to the public through the Library of Congress’ Chronicling America Collection database. This is especially noteworthy to the VTNP because once all items are digitized, the entire Vermont newspaper collection will be available to the world, which is a major shift from the microfilm format of preservation (Vermont Newspaper Project, 2011).

Concluding Remarks

As stated in the introduction, the digitization of newspapers is a large topic with much literature written on the topic. Scholars and researchers seem to agree when it comes to the advantages and disadvantages of digitization. With the creation of the National Digital Newspaper Program Standards for Digitization, institutions throughout the country can be a part of universally standardized newspaper digitization efforts. Since the digitization of newspapers is a new innovation in terms of access and preservation, more and more institutions are beginning to embrace the standards and make the shift from microfilm preservation with limited access, to digital preservation with unlimited global access for all.
References

http://library.state.ak.us/hist/newspaper/newspaper.html

http://chroniclingamerica.loc.gov/

Gore, Emily & Skinner, Katherine. (2010). The MetaArchive Cooperative: Chronicles in

Hasenay, Damir & Krtalic, Maja. (2010). Preservation of newspapers: Theoretical
approaches and practical achievements. Retrieved from
lis.sagepub.com/content/42/4/245.full.pdf

Illinois Newspaper Project at University of Illinois at Urbana-Champaign. (2009).

Illinois Newspaper Project at Chicago History Museum Research Center. (2011). Online

Microfilming. Retrieved from
http://www.alplm.org/library/library_depts.html#Newspaper

International Coalition of Newspapers. (2010). Call for Participation in the International
Coalition on Newspapers (ICON). Retrieved from
icon.crl.edu/PDF/ICON%20Participation.pdf

project). Retrieved from

International Newspaper Conference 2010: Digital Preservation and Access to


