

# National Endowment for the Humanities National Digital Newspaper Project Proposal

## PROJECT NARRATIVE

### Introduction

Founded in 1849 by the territorial legislature, the Minnesota Historical Society provides a variety of historical programs and services on Minnesota history. Although it receives generous support from the state, it is a private, nonprofit organization with a governing council elected by the membership. This public-private partnership has enabled the Society to grow from exhibition rooms in the former state capitol to one of the nation's largest historical organizations.

Among the Society's programs and services are a state historical museum; a statewide network of historic sites; educational services to schools across Minnesota; extensive collections on the history of the state and region; a major research library; the Minnesota Historical Society Press; the State Historic Preservation Office; State Archives; and technical assistance to local and county historical societies. The Historic Sites network, developed since 1958, ranges from Split Rock Lighthouse and Historic Fort Snelling to individual homes and archaeological sites. It is currently the largest historical society in the nation, with over 18,000 members, a staff of over 350 FTE's and an annual budget of approximately \$40,000,000.

The items in the Society's collections number in the millions and comprise the following categories: archaeology; three-dimensional; manuscripts; printed materials; art; sound and visual; oral history; and government records. The Library, Publications, and Collections Division has oversight of the collections. The Collections Department encompasses directly the responsibilities for collecting, managing, and caring for all the materials; the Collections Management Department provides the support functions of processing and conserving them. The Division provides extensive access to the collections through its library and especially through the use of technology and the World Wide Web.

The newspaper collection is one of the most valuable, most used and most extensive. The Society holds the largest collection of Minnesota newspapers of any repository, represented by over 4000 titles. The dates of the collection range from 1849 to the present and it includes daily, weekly, non-English language, labor, ethnic, reservation, legal, prison, religious, political and school papers. Virtually the entire collection has been microfilmed. Catalog records were created and contributed to CONSER/OCLC by the Society which was selected to represent Minnesota in the United States Newspaper Program (USNP). The newspaper collection receives intensive use from library patrons, for a variety of purposes, from scholarly research to family history, and serves users statewide and nationwide through Inter Library Loan (ILL). The collection is especially important because of Minnesota's role as a cultural and economic center in the Upper Midwest and its historical significance in the economic development of the entire Northwest.

Digitizing the newspapers and making them available online would greatly increase their usage and value. The Society has discussed such a project with a variety of entities, such as OCLC and Thomson-Gale, but has not yet moved forward with any such effort. In fact, no entity has done any substantive work involving the digitization of Minnesota's newspapers, with one exception: Winona State University has made available an online archive of three late 19<sup>th</sup> century and early

## National Endowment for the Humanities National Digital Newspaper Project Proposal

20th century Winona newspapers: the *Winona Argus*, *Daily Republican* and *Republican Herald*. This will eventually cover the period to 1925 and include over 123,000 pages of text.<sup>1</sup>

But the Winona newspaper digitization project is a tightly focused effort. As well, it has been done with microfilm supplied by the Minnesota Historical Society. Accordingly, given the breadth of its newspaper holdings and its participation in the USNP, its mission to serve the entire region, and its extensive experience with digitization and technology projects, the Society is best placed to work with the National Endowment for the Humanities and the Library of Congress to initiate such a program in Minnesota.

### History and scope of the project

Soon after the establishment of Minnesota as a territory, its first newspaper, the *Minnesota Pioneer*, was published on April 28, 1849 in St. Paul. The paper is still in publication today as the *St. Paul Pioneer Press*. Territorial Governor Alexander Ramsey noted the importance of newspapers as the “day books of history” in his message to the territory’s first legislative assembly and also noted the importance of preserving “a copy of each and every newspaper that may be published in the Territory.” The Society’s library has been collecting and preserving the territory’s and state’s newspapers ever since.

As Minnesota’s population increased, the publication of newspapers grew as well. Four daily newspapers were established in St. Paul in 1854, with growth reaching a peak of 45 in 1920. Weeklies reached their peak in 1910 with 641 newspapers in publication. Today the Society is collecting the 29 dailies and 398 weeklies published in Minnesota as well as special interest and neighborhood papers.

The Society has worked with newspaper publishers over the years to have as complete a collection of Minnesota newspapers as possible and many publishers have made copies of their files available for microfilming. In 1957, by Minnesota statute, all newspapers publishing legal notices have been required to file copies with the Society. This has helped greatly in maintaining as complete a record of the Minnesota press as possible.

Some notable examples of the newspapers that have been collected over the years include “foreign–language” papers reflecting the German, Swedish, Norwegian, Finnish, and French speaking and reading immigrants that settled in Minnesota in the late nineteenth and early twentieth centuries. Among other titles worth mentioning is the Minneapolis *Saturday Press*. Its actions resulted in the 1925 “gag law” and the consequent landmark lawsuit, *Near vs. Minnesota*. This case concerning freedom of the press eventually made its way to the U. S. Supreme Court, leading to in the 1931 decision supporting the constitutional guarantee of freedom of the press. The collection also includes the Stillwater *Prison Mirror*, which began publishing in 1887 and is the longest publishing prison newspaper in the United States. The Minneapolis *Appeal* (1889-1923) and the St. Paul *Recorder* (1934-2000) are long running newspapers representing the African-American interests of the Twin Cities area.

From the period 1880-1910, a variety of titles richly document the state and the region. These are all in the Society’s collections, microfilmed, but not digitized, nor available through private

## National Endowment for the Humanities National Digital Newspaper Project Proposal

sector sources such as ProQuest. From the capital city, there are the St. Paul *Dispatch*, 1868-1968; the St. Paul *Globe*, 1878-1905; and the St. Paul *News*, 1887-1938. From Minneapolis, there are the *Progress*, 1889-1978; *Register*, 1889-1923; and *Times*, 1890-1948. As well, there are substantial collections of regional, county and local newspapers that collectively represent the history of the state. All titles in the collections are described online, in a newspaper holdings database (at <http://collections.mnhs.org/newspapers/arsearch.html>) and in the library's catalog, described in standard MARC records (at <https://mnhs.mnpals.net>). Newspaper catalog records also are available in the OCLC national bibliographic database.

### Historical events, 1880-1910

The newspapers at the Society represent the history of all significant regional concerns, as Minnesota has long been the primary catalyst of cultural and economic activity in the Upper Midwest. In particular, Minnesota has set the pace for the regional economy, with the reach of its railroad, mining, milling and timber industries extending all the way to the Pacific. As well, the notable roles played by Native Americans and immigrant groups in the state set the framework of the understanding of major social, political and cultural trends in the 20<sup>th</sup> century.

Some specific events related to economic concerns to note are:

1875-1890: Bonanza farms, huge acreages created from the sale of land by the Northern Pacific Railroad to its investors to cover its debts, covered thousands of acres and produced large wheat crops. Through the creation of bonanza farms, Minnesota and North Dakota — and the Red River Valley in particular — became one of the country's largest wheat producing areas.

1880: Minnesota wheat and the power of St. Anthony Falls make Minneapolis the nation's capital of flour milling. A year later, Pillsbury's new A Mill is the largest flour mill in the world.

1882: The falls of St. Anthony in Minneapolis power the first hydroelectric central station operating in the United States.

1884: With the state's first shipment of ore from the Vermilion Range, Minnesota's iron industry is launched. Within 20 years, new immigrants will mine from the region a great majority of the iron for the nation's industrial boom.

1890: Duluth is booming on the promise of timber and mineral extraction. Its population has grown nearly ten-fold over the decade.

1891: George Hormel opens the Hormel meatpacking company at the right time. As corn replaces wheat in some southern Minnesota fields, it creates an abundance of hog feed and, as a result, a boom in hog farming and meatpacking

1891: German immigrant Frederick Weyerhaeuser, one of the most powerful men in American lumbering, moves his offices to St. Paul. Skilled at bringing

## National Endowment for the Humanities National Digital Newspaper Project Proposal

competitors together in huge undertakings, he makes heavy investments in Minnesota timber and mills before moving on to the Pacific Northwest.

1891: Conservationists win a bitter fight against lumber interests to establish Itasca, Minnesota's first state park, at the headwaters of the Mississippi River.

1893: James J. Hill pushes his Great Northern line to the Pacific Coast. The 1,816-mile track from St. Paul to Seattle completes the railroad he later calls his "great adventure."

1902-1903: The Minneapolis Millers and the Saint Paul Saints join six other midwestern baseball teams to form the American Association.

1908: Protesting low wages and a contract system that encourages workers to compete against each other, 8,000 iron miners walk off their jobs on the Mesabi Range.

Pertinent to Indians, these events are notable:

1884: The Court of Indian Offenses at Red Lake enforces rules forbidding plural marriages, dances, destruction of property following death, intoxication, liquor traffic, interference with the 'civilizing program,' and leaving the reservation without permission.

1889: The Nelson Act breaks up Ojibwe reservations into individual plots of land, leaving only Red Lake in tribal hands. The 160-acre family allotments are large enough for a farm, but too small to live on by hunting and fishing. The government sells leftover land to lumber companies.

1898: On 5 October, Anishinaabe from the Leech Lake Reservation fight U.S. troops at Sugar Point in the last Indian battle.

1907: Red Wing native Frances Densmore embarks on a life-long study of Indian music and culture. From a single recording of a performance by Kitchimakwa ("Great Bear") at White Earth, she eventually collects thousands of songs of the Ojibwe, Dakota, and 10 other tribes.

There are, in addition, a variety of miscellaneous events to note:

1880s: Mayos begin practice in Rochester

1885: St. Paul's *Western Appeal* begins publication for the city's growing black community.

## National Endowment for the Humanities National Digital Newspaper Project Proposal

1886: Dr. Martha Ripley opens a maternity hospital for unwed mothers in Minneapolis. A pioneer in combining social care with medical treatment, she taught new mothers how to care for their babies and helped them find work.

1891: Headwaters of the Mississippi River becomes Minnesota's first state park, as conservationists win a bitter fight against lumber interests.

1894: More than 400 people die in the Hinckley forest fire. The scattered remains of harvested trees are kindling for the "red demon" that destroys 900 square miles in only a few hours.

1898: Olaf Ohman turned up a stone on his farm near Alexandria and sparked a controversy. What appears to be ancient Norse writing on the stone indicates that Viking explorers reached Minnesota in 1362, 130 years before Columbus' voyage.

1908: Louis Sullivan's National Farmers' Bank opened in Owatonna. It is perhaps the most famous small-town bank in the country.

### Methodology and standards

The Society's newspaper microfilm collection currently consists of 68,956 rolls. Of these, 10,159 rolls represent the period from 1880-1910. A number of these rolls would be excluded from consideration due to copyright issues, but the collection would still have around 7500 rolls of film from which to select for the project. At an average of 500 images per roll, there are around 3,750,000 available images.

Some historical background will demonstrate the quality of the film. In 1948 the Society began a preservation microfilming program. Until that time the Society's newspapers had been routinely bound and maintained in hard copy form. Space to store the volumes was rapidly diminishing, so the microfilming program was born.

From 1948 through 1961 the microfilm operation created only master negatives, which were used not only as preservation copies, but also as reference copies. The film created during this period was of inconsistent quality for a variety of reasons, two of the most noteworthy being that available space at the time located the cameras in a room with a large window, making it difficult to maintain consistent lighting, and the raw film stock emulsions lacked the consistency of what is available in the present. Almost all of the film produced during this time was of contemporary newspapers.

In 1961 a new microfilm lab manager made significant improvements in the operation. The existing negatives were removed from public use and vesicular film copies were made for reference. A stringent quality control program was instituted, very much like the one in current use. In 1978 the Society began an ambitious 10-year program to microfilm its collection of bound newspaper volumes. As the project progressed, it borrowed newspapers from other repositories to fill gaps in its collection. In 1989 the Society received an NEH grant to

## **National Endowment for the Humanities National Digital Newspaper Project Proposal**

participate in the USNP in order to complete the backlog. The grant provided funding to film the volumes for the years from 1920 through 1945.

Presently the MHS microfilm lab staff films approximately 400 titles of current Minnesota newspapers. Filming is done according to ANSI/AIIM standards for newspaper microfilm. Prior to microfilming, the papers are collated into logical bibliographic units and flattened with a common household iron. The master negatives are created on Kodak MRD-2 cameras on 35mm polyester base microfilm. The film is processed in an Allen Products M-30B deep tank processor. Methylene blue tests are performed each day film is processed.

The microfilm project director performs a preliminary quality check of the processed film by examining the filmed resolution chart under a microscope, measuring the density of the gray control card and several images throughout the roll, and examining the roll for gross defects such as scratches, fogging or other noticeable blemishes. Microfilm staff members then inspect each bibliographic roll frame by frame on a microfilm reader. Any errors are noted and the applicable pages are refiled and spliced into the original roll using ultrasonic splices. After the film has passed the final inspection a silver duplicate negative is created and a vesicular positive copy is made for reference use. The security (master) negative is stored offsite at Iron Mountain in Pennsylvania and the working (duplicate) negative is stored in an environmentally controlled vault at the Minnesota History Center.

The microfilm lab produced a large majority of the microfilms for the period 1880-1910 during the years from 1978-1982. Resolution and density of the film produced in this period are generally quite good and meet applicable standards. Microfilm produced from 1962 through 1977 and from 1983 to the present also meets these standards. The lab used acetate base film stock until late in 1984, which was what was available at the time

In a digitization project, the Society will follow the metadata specifications promulgated by the Library of Congress. The Society produces high-quality, full-level bibliographic records for all newspaper titles in its collections, each of which is available in the CONSER/OCLC catalog. The catalog data will be the basis for the descriptive metadata for each newspaper title. Structural metadata (e.g. date, edition, sequence information, page number) will be provided for each page, issue, and title. Technical metadata will be provided for each page scan. Additional technical metadata describing the quality parameters of the film negative will be provided for each microfilm reel that is digitized.

### **Selection process**

As the principal criteria for selecting newspapers for digitization, the Society would follow the direction set by the NDNF, identifying the titles that reflect the political, economic and cultural history of the state; which provide broad coverage of the state and its population; and which have a broader chronological span.

There is no standard reference work on the history of newspapers in Minnesota, so there is no directly pertinent intellectual work from which to draw for analysis. There is, however, a context of references and articles from *Minnesota History*, the journal of the Society. This dates from

## **National Endowment for the Humanities National Digital Newspaper Project Proposal**

1915 to the present day; its index is available online and the Society is in the process of digitizing all the back issues. From this resource, along with the materials in the library, the Society's reference staff will be able to compile a briefing book for the advisory board, with lists of the significant events of the period, 1880-1910; background on the prospective choices of newspapers for digitization; and sample articles from each. This briefing book will form the raw material for the essays on the three decades and on the selected, digitized newspapers that the Society will submit to the Library of Congress.

The advisory board is comprised of a diverse set of constituents, representing researchers, libraries and archives. The members are identified in Attachment 7. The board will meet twice each year. In the first year, the board will meet: 1) to review the prospective titles and make a preliminary selection; and 2) to review the digital samples from the selected titles and confirm the selection. In the second year, the board will meet twice as well, first to review the progress of the project and then to review and approve the final products of the effort.

The Society and the board will work with staff from the National Endowment for the Humanities and the Library of Congress in the selection process. It will also consult with libraries already engaged in the NDNP, to understand better their efforts and how they might inform work in Minnesota.

As noted, the initial focus will be on newspapers published in the Twin Cities, as these will have the broadest coverage, both in terms of content and geography. They are also published in English. The project will not consider titles which are still in production, but it will take a closer look at those titles that have production runs throughout the period covered by the NDNP and beyond. The Society holds the master negatives to all these; the records indicate as well that these are, if not perfectly complete, certainly the most complete that are available. Other microfilm versions came from the Society.

### **Infrastructure for digital projects**

The Society has extensive experience with digital content and technology projects. As the most recent and most notable recognition of that, the Society received a grant from a local foundation to expand its technological infrastructure. This builds on a variety of efforts that have demonstrated a broad capacity for collaboration and innovation in the areas of digitization, digital content and technology.

In 2006, the Bush Foundation awarded the Society \$997,000 for a three year project, as part of the Foundation's Large Cultural Organization Development Fund to expand its technological infrastructure. With that additional capacity, the Society will be able to support:

- build a regional network that will integrate access to the collections of diverse institutions and facilitate resource discovery;
- enhance its capabilities through more robust, standards based, integrated systems that utilize best of breed solutions;
- increase the quality and quantity of available digital content, in partnership with a variety of institutions;

## National Endowment for the Humanities National Digital Newspaper Project Proposal

- sustain the infrastructure through the use of solutions based on open standards; and
- ensure the necessary financial base by creating a demand for valued products that will result in ongoing funding and by offering particular goods and services that will provide direct financial underwriting.

The partners in the project are the states of North Dakota and South Dakota, along with four Minnesota organizations: the Anoka, Blue Earth, Nicollet and Olmsted County Historical Societies. As the first step, the Society has recently concluded negotiations to purchase and implement a federated database application, from Autonomy Inc.

This project and the application build on an extensive array of digital content developed by the Society. The primary completed projects include:

- *Visual Resources Database*, with an index of 189,000 records and the associated images of 120,000 photographs and art works;
- *PALS*, the online library catalog, the guide to over 175,000 books and close to 100,000 cubic feet of manuscripts and government records;
- *Collection Management System*, a KE EMu application with descriptions for over 230,000 items in the museum collections;
- *Name indices*, over 20,000 place names, a projected 1.75 million birth records and 2 million death records;
- *Minnesota Greatest Generation* reminiscences, with a planned 1000 stories; and
- close to 1000 EAD (Encoded Archival Description) formatted finding aids.<sup>2</sup>

Other projects already underway include:

- *digitization of vital records* – to provide access to 1.75 million birth records online, with funding from the Minnesota Department of Health;<sup>3</sup>
- *web interface to the collections management system* – to increase and enhance access to the Society’s database on its collections, principally three-dimensional and archaeological, already part of the Library Division FY2007 budget;
- *web-based geospatial data and applications* – to package digital content for a specific audiences, in collaboration with the Society’s education programs and the state’s K-12 teachers in support of Minnesota’s history and geography standards, currently supported by a project funded by the Institute for Museum and Library Services;<sup>4</sup>
- *preservation of legislative digital records* - a collaboration with the Minnesota and California Legislatures, along with the San Diego Supercomputer Center, funded by the National Historical Publications and Records Commission;<sup>5</sup> and
- *federated/distributed database access* - purchase of Autonomy’s IDOL search engine, allowing for single interface to multiple databases (implementation expected in early 2007).

The Society has not had extensive experience in creating digital content from microfilm, but it has intensively researched the process in preparation for this proposal. Most of its conversion work has been done from originals – e.g., the 1.75 million birth records, done in-house, or the

## National Endowment for the Humanities National Digital Newspaper Project Proposal

some 3000 maps in Minnesota Maps Online, done, with vendors. The Society is currently planning to convert the back issues of its journal, *Minnesota History*. In preparation for that effort, staff has investigated vendors, standards, practices and models. Much of this has been codified in an in-house manual. The Society is now selecting a vendor to digitize the material and plans to begin work by the end of the calendar year. At present, Kirtas Technologies is the most likely contractor, although the agreement has not yet been finalized.

As a collaborator, the Society has been, with the University of Minnesota, one of the lead partners in the Minnesota Digital Library Coalition, a partnership supported by LSTA funds to digitize materials documenting the cultural heritage of the state. The project's principal goal has been to engage and educate repositories across Minnesota in the development of digital content. *Minnesota Reflections* is the initial digitization effort of the MDLC. This digitization project, conducted from 2004-05, involved more than 50 participating historical societies, special archives, and libraries. The MDLC and participants digitized more than 6,000 unique photographs and images. Another phase of digitization is just now underway.

Finally, the Society has extensive experience working with XML, especially in its relation to metadata. It has long been a leader in the application of encoded archival description (EAD) and is working with encoded archival context (EAC). The staff is extremely familiar with the processes of aggregating and converting metadata in different formats.

Because of these experiences, the Society has an experienced staff and a notable reputation for innovation and achievement in the application of technology. It has collaborated successfully with a number of institutions, both locally and nationally. It has been a leader in a number of technological areas, particularly in the areas of electronic records and metadata.

### **Work plan**

In general, the Society will adhere faithfully to technical guidelines and standards of the NDNP and the Library of Congress; these create the framework for developing the work plan. In particular, the Society proposes to structure the project in several phases, with the actual digitization outsourced, but with the critical roles of preparation, metadata management, quality control and transfer to the Library of Congress handled in house. Staff of the Society, along with the project advisory board, will manage those responsibilities.

### **Selection of newspapers**

As described more fully above, the Society has received the commitments of a diverse and talented group for the project advisory board; together, they represent the significant constituencies for digital content such as this. Working with the Society's staff, they will determine which newspapers are optimal for digitization.

The Society already has a good grasp of the physical quality of its microfilm, due to its primary role in all phases of the microfilming process. Essentially, the Society has managed every aspect of the collection, from acquisition to providing access. But, following the preliminary selection of newspapers, Society staff will examine more closely each roll, to insure that the subsequent

## **National Endowment for the Humanities National Digital Newspaper Project Proposal**

digitization would run smoothly. This examination would address the quality of the film as well as the completeness of the microfilmed run, to validate what the bibliographic record indicates.

At this point, the Society will have samples of the film digitized, to allow the staff and advisory board the opportunity to verify the initial assessment. If the review is positive, the Society staff will create duplicate negatives and ship them to the vendor for conversion to digital format.

### **Scanning and OCR**

The actual digitization of the selected newspapers will be outsourced. Because of institutional policies on contracting for large scale projects, the Society will have to produce a formal request for proposals and evaluate the responses before choosing a vendor. Accordingly, we cannot give precise details on the cost or the contract at this time; we are limited to providing the information from informal and non-binding discussions with potential vendors.

Society staff contacted several potential vendors for information. All had substantial experience digitizing from microfilm; some were currently working on NDNP projects. Each was confident of its ability to manage a contract with the Society. They provided a range of quotations from approximately \$.85 to \$1.60 per page for digitization, including scanning, optical character recognition and the creation of structural and technical metadata.<sup>6</sup> The final selection of a vendor will not be solely dependent on price; the Society can take into consideration a variety of factors to determine the best offer, rather than simply accept the lowest cost. For the purposes of this application, the budget reflects the higher estimate; if the selected vendor charged a lower rate, the Society could either digitize additional pages or simply not request the unnecessary funding from the NEH.

The specifications for the scanning process will require the vendor to:

- scan from a clean second-generation duplicate silver negative microfilm (produced by the Society and then to be deposited with the Library of Congress);
- capture images at 8-bit grayscale at a resolution of 400 dpi, if possible; otherwise, 300 dpi (relative to the physical dimensions of the original newspaper, rather than the microfilm);
- scan a standards-based target film strip at the start of each session, to monitor scanning equipment performance. Target test images should be delivered along with the page images;
- provide the master page images, delivered to LC, as uncompressed images in TIFF 6.0 format.

For each page of the newspaper, the vendor will reproduce the image in two raster formats, grayscale, generally with a resolution of 400 dpi, if possible, otherwise, 300 dpi, in an uncompressed TIFF 6.0 format, and the same image, compressed as a JPEG2000 file. The vendor will also create a file with OCR text and associated bounding boxes for words, 1 file per page image, along with a PDF image with hidden text, text and image correlated.

Per the project guidelines, the OCR will meet these specifications:

## National Endowment for the Humanities National Digital Newspaper Project Proposal

- one OCR text file per page image. (Discrete files should be produced for each page, rather than for a multi-page issue or entire title);
- each OCR text file name corresponds to the page image it represents;
- text in UTF-8 character set;
- no graphic elements saved with the OCR text;
- OCR text ordered column-by-column (that is, in a natural reading order);
- OCR text file with bounding-box coordinate data at the word level;
- OCR will conform to the ALTO XML schema.

If possible, the vendor will provide these additional elements for the OCR files: confidence level data at the page, line, character, and/or word level; and point size and font data at the character or word level. In sum, a total of 100,000 OCR files and corresponding images will be created.

### **Metadata**

The Society will follow the metadata specifications determined by the Library of Congress. Some metadata will be produced by the digitization vendor and some by the Society's staff, but all the metadata will be aggregated and evaluated by the Society staff, for packaging in a METS object structure and delivery to the Library of Congress.

Standard descriptive metadata for the Society's newspapers holdings already exist, in both MARC bibliographic records and the Society's newspaper database. Appendix 4 contains a description of the Society's practices for cataloging newspapers in bibliographic records, as well as a data model for the newspaper database. Additionally, the Society's MARC catalog records will be reviewed, updated as needed, and holdings confirmed in the CONSER/USNP Union List prior to submission of records and metadata to the NDNP. The descriptive metadata will follow an extension schema such as MODS based on the MARC record and holdings data and will be encoded in XML.

Not all the technical metadata the Society has compiled on the microfilming process is pertinent. What is pertinent will be extracted and will be combined with technical metadata on the scanning and OCR process, created by the vendor. The Society's staff will supplement this data with additional technical metadata to support the functions of a trusted repository. The vendor will also provide structural metadata for pages, issues, editions, and titles to support a chronologically based browsing interface, and to identify image and OCR files.

Staff will also perform standard quality control evaluations to ensure that all metadata meets project specifications. This will be modeled on the quality control process the Society developed for its birth record scanning effort, which involves an image by image review. Metadata will be validated as directed by the NDNP and using the NDNP provided software. The digital asset metadata will be delivered in XML batch files using METS object structure.

## National Endowment for the Humanities National Digital Newspaper Project Proposal

### Transfer

In the course of the project, the Society will complete all reports, as required to the NEH. It will transfer these products to the Library of Congress:

- decade essays: the history of state's newspapers and impact on regional historic events for the relevant time period, one essay per decade, with a maximum of 1,000 words per essay; and
- newspaper essays: describing the scope and content of each title, with its history and significance, with a maximum of 500 words per essay.

The briefing book will form the raw material for the essays on the three decades and on the selected, digitized newspapers that the Society will submit to the Library of Congress. For each microfilm reel digitized, the Society will transfer a second-generation duplicate silver negative microfilm, made from the camera master, barcoded (LC to supply barcodes for all reels).

For the digital content, the Society will transfer:

- technical metadata concerning the quality characteristics of the film used for digitization, encoded in a METS object with other digital assets;
- validated master digital page image format = TIFF 6.0 uncompressed;
- validated OCR text file with bounding-box coordinates = 1 text file per page;
- validated PDF Image with Hidden Text = 1 PDF per page;
- validated derivative digital page image format = JPEG2000 (.JP2) using specified compression options; and
- validated metadata using METS in accordance with guidelines.

### Schedule

#### 1. July – September 2007

Develop RFP and specifications for scanning vendor  
Work with Library of Congress and NEH; attend program meeting at Library of Congress  
Schedule advisory board meeting  
Research and write briefing book for advisory board

#### 2. October – December 2007

Hold advisory board meeting  
Select newspaper titles  
Review quality of microfilm  
Select vendor and negotiate contract  
Review existing metadata  
Submit semi-annual project report  
Duplicate microfilm for scanning

#### 3. January – March 2008

**National Endowment for the Humanities  
National Digital Newspaper Project Proposal**

Transfer sample of microfilm to vendor for scanning and OCR  
Establish quality control procedures for review of images and metadata  
Continue review of microfilm and existing metadata  
Develop METS specifications

4. April – June 2008

Hold advisory board meeting  
Review digitized sample; perform quality control and confirm quality control process  
Review and aggregate metadata; confirm metadata development process  
Submit semi-annual project report  
Standardize procedures for routine transfer of microfilm to vendor

5. July – September 2008

Send second batch (in quantity to be determined) of microfilm to vendor  
Perform content and metadata quality control  
Standardize procedures for transfer of images and metadata to Library of Congress  
Attend program meeting at Library of Congress

6. October – December 2008

Hold advisory board meeting  
Submit semi-annual project report  
Transfer images and metadata on monthly basis to Library of Congress  
Perform content and metadata quality control  
Research historical and descriptive essays

8. January – March 2009

Transfer images and metadata on monthly basis to Library of Congress  
Perform content and metadata quality control  
Draft historical and descriptive essays  
Transfer final set of images and metadata to Library of Congress

9. April – June 2009

Hold advisory board meeting  
Duplicate microfilm for Library of Congress  
Submit duplicate microfilm to Library of Congress  
Submit final versions of historical and descriptive essays  
Submit final project report

## National Endowment for the Humanities National Digital Newspaper Project Proposal

### Staff

*Robert Horton, Project director.* Horton will direct the project, with responsibility for the budget, convening and facilitating the advisory board, routinely meeting with the managers and overseeing the development of semi-annual and final reports. He is the director of the Library, Publications and Collections Division of the Society, in which capacity he has developed and managed numerous large scale technology projects. Horton will spend 5% of his time on the digital newspaper project.

*Irene van Bavel, Manager.* Will be responsible for the overall management of the project. Her responsibilities will include liaising and communicating with funding agencies, the Library of Congress, and contractors. She will develop a detailed project plan, coordinate the work of cataloguers, preservation staff, metadata creators, and digitization staff, and provide regular project updates. She will be directly responsible for outsourcing the imaging and text conversion for this project and will develop digitization guidelines, select a digitization vendor, establish schedules and deliverables, and ensure quality control of the material. Ms. van Bavel is Manager of Digitization Projects at the Society. She has extensive experience in project management and digital content production and has undertaken numerous large digital conversion projects involving digital content from microfilm. She will spend 15% of her time on the project.

*Steve Cunat, Microfilm manager.* Cunat is the microfilm project manager at the Society and will be responsible evaluating the physical properties of the film of the newspapers selected for digitization, duplicating the film and shipping it to both the vendors and the Library of Congress. He will also be involved in the quality control review of the digitized images and the newspaper metadata. Given his long experience working with newspapers in the project, he will play a role in the project management and planning process. He will devote 10% of his time to the project.

*Anne Levin, Metadata and quality control coordinator.* Levin is an experienced MARC21 cataloger with expertise in archival, monographic, and serials cataloging, will be responsible for developing and updating content metadata, and for aggregating descriptive, technical, and administrative metadata into the METS wrapper. Anne routinely creates original MARC21 catalog records and Encoded Archival Description finding aids for manuscript collections; creates original, and maintains existing, MARC21 catalog records for newspapers using CONSER and USNP guidelines; creates catalog records in local system (Aleph) and OCLC using standard cataloging tools; and provides NACO authority control for bibliographic records. She will also supervise the quality control process for both images and metadata. She will spend 25% of her time on the project

*Kathryn Otto.* Otto is head of the Reference Department at the Society's Library. She will supervise the work of the Reference Department staff on the development of the briefing book for the advisory board, supporting the selection of newspapers, and then the development of the historical and evaluative essays required among the project's deliverables. She will also participate in the project's management and planning process. We estimate the staff of the department will spend 40 hours per year on research and writing. Otto will devote 5% of her time to the project.

**National Endowment for the Humanities  
National Digital Newspaper Project Proposal**

*Scanning Technician, Quality Control.* Staff from the Collections Management Department will ensure quality of digital images and text conversion from the digitization vendor. The staff will examine a subset of digitized documents and text outputs from the vendor at the beginning of the project. Once the digitization and OCR text is completed according to set guidelines, the scanning technician will review the adherence to established requirements, quality of the images and accuracy of uncorrected OCR text, structural and technical metadata. The staff time will equal 250 hours per year, an estimate based on the Society's experience with its project to digitize the state's birth records.

*Advisory board.* Members include Annette Atkins, Professor of History, College of St. Benedict/St. John's university; Brenda Child, associate Professor, American Studies Program, University of Minnesota; Donna Gabaccia, Professor of History and Director of the Immigration History Research Center, University of Minnesota; Mark Geiger, Population Center, University of Minnesota; Beth, Kaplan, University Archivist, University of Minnesota; Joe Mount, Winona State University Library; and Greg Simpson, St. Paul Public Library.

**National Endowment for the Humanities  
National Digital Newspaper Project Proposal**

**Endnotes**

<sup>1</sup> For details, see the project web page: <http://www.winona.edu/library/databases/winonanewspaperproject.htm>.

<sup>2</sup> A more complete list is available in Appendix B.

<sup>3</sup> The Society makes the records available online as they are scanned and indexed. The site's address is <http://people.mnhs.org/bci>.

<sup>4</sup> The project is described online at <http://www.mnhs.org/preserve/records/geographyonline/main.htm>.

<sup>5</sup> For more information, visit the project web site:

<http://www.mnhs.org/preserve/records/elegislature/elegislature.htm>.

<sup>6</sup> Copies of sample quotations are attached in Appendix 5.