The New York Public Library

National Digital Newspaper Project
Creating Digital Access to Significant New York State Newspapers, 1880-1910
July 2007-June 2009

Project Description

The New York Public Library (NYPL) seeks $301,985 to continue its participation in the National Digital Newspaper Project (NDNP), partnering with the National Endowment for the Humanities and the Library of Congress to create digital access to historical newspapers. Building on its participation in the preliminary phase of the NDNP project, the Library proposes to digitize a minimum of 100,000 pages of *The Sun* and the *Evening World* published from 1880 to 1910.

New York's lively press has borne witness to a rich tapestry of historical events. The New York Public Library's holdings of State newspapers, in print and on microform, are extensive, with many titles that were collected solely by the Library or by only a few other libraries. The New York Public Library holds the master negative microfilm for these titles, which ensures access to the source materials, and has identified other institutions that also hold these titles on microfilm, which will help the Library fill any gaps in the title runs, if necessary.

The Library will work with a digitization vendor in order to take advantage of the best available equipment. The project will create digital images in TIFF 6.0, PDF, and JPEG200 formats and OCR text files for a minimum of 100,000 newspaper pages. Technical and structural metadata and XML packets will be produced for every image, as well. Library staff will select and prepare the microfilm materials, organize and track the digital elements for each image, and perform quality control for the materials that are produced.

The Library will contribute the efforts of staff in its Digital Library Program, the Goldsmith Preservation Laboratory, and the Humanities and Social Sciences Library, one of the four NYPL Research Libraries. The total project budget is $382,841, with $301,985 requested from the Endowment and $80,856 contributed by the Library.

The New York Public Library
National Digital Newspaper Project

Creating Digital Access to Significant New York State Newspapers, 1880-1910
July 2007-June 2009

Project Narrative

Since it was founded in 1895, The New York Public Library (NYPL) has been committed to building and preserving its distinctive collections and providing free and open access to these resources for the benefit of all who come in search of them. The Library's collections of newspapers are among its most valuable and heavily used holdings. Large and frequently unique backfiles constitute a rich mine of data and opinion for the historian as well as the general researcher. Many of the Library's newspaper collections have been microfilmed for posterity.

While microfilm is the current preservation standard, it also presents a number of limitations. Transferring content from microfilm to the digital medium offers numerous benefits to the researcher. Microfilm must be used onsite, but digital content may be accessed remotely through any online connection. Scanned materials may be made searchable, a function that does not exist with microfilm. In addition, unlike with the microfilm format, the quality of the online materials is unaffected by the number of times they are accessed.

With support from the National Endowment for the Humanities (NEH), The New York Public Library is currently digitizing 100,253 pages of the historically significant New York State newspapers *The Sun* and the *Evening World* as part of the initial phase of the National Digital Newspaper Project (covering the years 1900-1910). As the National Endowment for the Humanities and the Library of Congress (LC) continue in their effort to create an online database of historical newspapers, The New York Public Library looks forward to contributing not only more of its extensive holdings but also its technical expertise in preservation and creation of digital archives. The Library requests a grant of $301,985 to participate in the second phase of the National Digital Newspaper Project, to make an expanded range of issues from *The Sun* and the *Evening World*, covering the years 1880-1910, available to the global community of researchers and help create a lasting resource for the future.

History and Scope of the Project

History of New York State Newspapers

The first newspaper published in the province of New York was the *New-York Gazette*, founded in 1725. Its publisher, William Bradford, was the official government printer for the province and thoroughly royalist in his politics. New York was the third colony to have a newspaper, after

Massachusetts and Pennsylvania. Bradford's former apprentice John Peter Zenger founded the second newspaper in the colony, the *New-York Weekly Journal,* in 1833. In contrast to Bradford, Zenger was highly critical of the colonial administration. His efforts to expose malfeasance were famously rewarded with arrest and trial for seditious libel. Zenger's attorney argued that the truth of what was written should be a justifiable defense in a libel trial—contrary to the law in force. Zenger was acquitted and remains a symbol of the freedom of the press.

By the time of the American Revolution, 22 newspapers had appeared in New York province, and all but one of them was published in New York City. The contents of the newspapers in these years consisted mainly of shipping news, stories reprinted from London papers, advertisements, official government notices, and essays that had appeared in British literary publications. After the Revolution, a lively press emerged with links to the main political factions—later political parties—of the day. The *Daily Advertiser*, which had the distinction of being the first newspaper to publish every weekday, took a staunchly Federalist position on the adoption of the Constitution and new Federalist papers appeared upstate. In 1801, Alexander Hamilton founded the *New York Evening Post* to further the interests of the Federalist party. The Anti-Federalists or Jeffersonians founded their own newspapers as a response.

By 1825, most New York State towns of reasonable size had a newspaper. Printing innovations and improved transportation led to a dramatic rise in circulation in the 1830s. The rise of the "penny" press transformed newspaper publishing. These newspapers were generally politically independent and appealed to the urban working and middle classes, not the elites. The first successful penny paper was the *Sun*, established in New York City by Benjamin Day in 1833. It was quickly followed by imitators. By 1840, New York State had more than 200 newspapers; by 1855, it had 671 newspapers, and by 1860, every major city had at least one daily. New York City was the center of newspaper publication, not just for the state of New York, but also for the entire country. As the center of U.S. trade—by 1860, two-thirds of all goods imported to the U.S. passed through the Port of New York—the City had become the hub for the dissemination of foreign and domestic news. New York City picked up the latest European news, and it then spread along the trade routes. Park Row in downtown Manhattan became known as Newspaper Row because it was the location of the offices of so many of the City's dailies.

The New York *Herald* in 1860 had a circulation of 77,000 and was the world's largest daily. During the years prior to the Civil War, newspapers played an important role in disseminating the issues. After the War had begun, newspapers were crucial in connecting the homefront with firsthand accounts of battles, profiles of military personalities, and political news. In New York City, where the war was not universally popular, newspapers were generally restrained in their support. While Horace Greeley's *Tribune* was thoroughly abolitionist, James Gordon Bennett's *Herald* supported Lincoln without enthusiasm. In the 1864 election, only five of the 17 dailies strongly endorsed Lincoln's re-election. The draft riots of the previous summer had demonstrated just how unpopular the War was among some of the City's inhabitants.

As the population grew, so did the circulation of the newspapers, not least because of the dramatic increase in immigration in the latter half of the 19th century. To appeal to various ethnic communities, a number of specialized foreign language titles appeared. The earliest (1828) was French, soon followed by titles in German (1834), Italian (1850), Spanish (1870), Yiddish

(1876), Czech (1882), Chinese (1883), Arabic (1892), Greek (1894), Croatian (1898), Hungarian and Slovak (1899), Russian (1910), Serbian (1911), Polish (1913), and Ukrainian (1919). But the increasing use of graphics also made the general interest newspapers appealing to immigrants and the working classes. Color printing was first used by the New York *World* in 1893 and color comics became a feature of the Sunday edition. Not long after, William Randolph Hearst purchased the *New York Journal* and, in competition with the *World*, gave rise to so-called yellow journalism, marked by sensationalism and loose treatment of the facts. Even so, such excellent writers as Nelly Bly, Jacob Riis, Stephen Crane, and muckraker Lincoln Steffins worked as reporters on New York dailies. *The New York Times*, founded in 1851, had a renaissance when it was taken over by Adolph S. Ochs in 1894. By the time of World War I, it had attained great prestige and reputation as the most dependable source of news in New York.

After the war, tabloid newspapers featuring prominent front-page photographs and sensational headlines appeared. Their daily diet of sports, scandal, and lurid crimes proved popular among the working classes, their format perfect for reading in a crowded subway car. Before the start of World War II, the *Daily News* had achieved a circulation of 2 million. Competition from the tabloids and the effects of the Great Depression combined to bring about the mergers of many newspapers and the loss of others to bankruptcy. The *Herald* merged with the *Tribune* in 1924, and the *World* with the *Telegram* in 1931. In 1950, the *Sun* merged with the latter to form the *World-Telegram and Sun*. Strikes against newspapers by reporter and trades unions in the 1960s dealt fatal blows to five New York City dailies, leaving just three survivors.

Although 98 percent of U.S. cities now have only one newspaper, New York City stands in contrast. Depending on what is counted, New York now has between three and six English language daily newspapers: *New York Times, New York Post, Daily News, Newsday, Wall Street Journal*, and *The Sun* (this title an homage to the penny paper of the past). Add to these approximately 200 ethnic and community based titles. Part of the explanation for this large number of choices is that New York is the major media market in the United States. However, another interesting theory is that it is also because New York has the greatest number of commuters using public transportation, which leaves the hands free to hold a newspaper. Subway riders are now greeted daily by the modern day equivalent of newsboys distributing two free dailies, *AM New York* and *Metro*. Whether the latter tabloids—designed to appeal to younger readers who do not read the traditional dailies—result in forming the next generation of serious daily newspaper readers, or instead hasten the decline of the traditional newspaper remains to be seen.

Historical Highlights

New York's newspapers have borne witness to a rich tapestry of historical events. Highlights of important political, social, and cultural news that occurred within the 1836-1922 time frame of the National Digital Newspaper Project, with a particular emphasis on the decades of 1880-1910, follow.

In 1836, Martin Van Buren, born near Albany of Dutch stock, became the first of several New Yorkers to be elected President of the United States. New York State was rapidly industrializing and by mid-century, it was the leading industrial state in the nation. New York City played a

dominant role in national commerce, including trade in southern cotton. By the eve of the Civil War, there were more than 4 million residents in New York, making it the most populous state. In 1860, New York City reached a population of 1 million, the first U.S. city to do so.

Much of the population growth in New York's cities was a result of the waves of immigrants coming to the U.S. to seek a better life. Beginning in the 1840s, Irish famine immigrants and Germans fleeing repression after the Revolutions of 1848 became the dominant new groups. A Nativist reaction resulted in the creation of the No Nothing Party. The Astor Place Riot of 1849, ostensibly a fight between partisans of the actors Macready and Forrest, was really the result of Native vs. Irish tensions.

Although no longer a slave holding state—New York had phased in abolition by 1827—the issue continued to cause controversies within the State's political parties. Some members supported full abolition and others merely opposed the extension of slavery to the territories. The Kansas-Nebraska Act intensified divisions, and ultimately resulted in the demise of the Whig Party and the formation of the State's Republican Party. In 1856, with strong support upstate and in the west, Republican presidential candidate John C. Fremont carried New York State and John A. King was elected the first Republican governor of New York.

Four years later, Abraham Lincoln carried New York easily for the Republicans, but captured only one-third of New York City's vote. When Civil War came, New Yorkers gave it lukewarm support. The Democrats opposed secession—although New York City Mayor Fernando Wood suggested that the City secede—but wanted a political settlement and were quick to criticize Lincoln for abuse of executive power. New York newspapers reflected these tensions, as for example, the *New York Tribune* supported a Radical Republican line, and the *New York Journal of Commerce* that of the Copperheads. The introduction of the unpopular military draft resulted in the bloody draft riots of July 1863 when a mob of 50,000 people terrorized New York City for three days. Beatings, lynchings, and arson went unchecked until Lincoln deployed troops.

After the Civil War, there was tremendous economic growth. The engineers of this growth were either captains of industry or robber barons depending on one's point of view. The increasing wealth of the elites brought about "The Gilded Age," and New York City was its epicenter. Ostentatious displays of wealth were a symbol of this period. In 1892, 27 percent of all American millionaires lived in New York City. At the same time, the wretched conditions of the laboring classes resulted in the growth of trade unions and labor agitations. Political reformers battled the corruption of the Tweed Ring of the Tammany Organization. "Boss" Tweed, symbol of corruption, went to jail for stealing millions from the public treasury. His nemesis, reform Governor Samuel J. Tilden, went on to be denied the Presidency of the U.S., although he had a majority of the popular vote. The Compromise of 1877 made Republican Rutherford B. Hayes President in return for the withdrawal of federal troops from the South, effectively ending Reconstruction. In 1895, The New York Public Library was founded, in part from Tilden's bequest. New York Governor Grover Cleveland had better luck than Tilden and was elected President in 1884, defeated for a second term in 1888, but re-elected in 1892.

New York State grew by 40 percent between 1880 and 1900. By 1880, rural areas of the state had begun to lose population, and by 1900, 56 percent of all New Yorkers lived in larger cities.

In 1883, the Great East River or Brooklyn Bridge linked New York City and Brooklyn, the two largest cities in the U.S. Much of the steady population increase was a result of increased immigration. France presented the U.S. with the Statue of Liberty in 1885, and Joseph Pulitzer organized a campaign in *The World* to raise funds to finance creation of a pedestal. In 1892, the Ellis Island Immigration Station opened in New York Harbor, and immigrants from Southern and Eastern Europe dominated. Immigrants settled in cities and lived in miserable conditions. Jacob Riis published *How the Other Half Lives* in 1890 and opened the eyes of many to the state of society's less fortunate. Social reformers opened settlement houses and fought for "New Law" or dumbbell tenements that provided for better air circulation and healthier living conditions. In 1898, Greater New York was created from the consolidation of the boroughs of New York, Bronx, Brooklyn, Queens, and Richmond.

William Randolph Hearst, publisher of the *New York Journal* and the *Evening Journal* used sensationalism to sell newspapers. He and Joseph Pulitzer embraced "yellow journalism" and contributed to the atmosphere that resulted in the 1898 War with Spain. When the war ended, the United States became a superpower with overseas territories. It was widely believed that an anti-McKinley *New York Journal* editorial inspired anarchist Leon Czolgosz to assassinate him in September 1901. Upon his death, Vice President Theodore Roosevelt, former Governor of New York, assumed the Presidency.

The decade 1900-1910 was defined by rapid urbanization, industrialization, population growth, and mobility that generated significant social turmoil. New York State newspapers both chronicled and reflected that turmoil. In the following years, New York newspapers would cover a wide range of political, social, and cultural events, both international and domestic, including: the death of Queen Victoria, the organization of baseball's American League in New York, the Russo-Japanese War, the Russian Revolution of 1905, the opening of the New York City subway system, the sensational trial of Harry K. Thaw for the murder of Stanford White, the founding of the National Association for the Advancement of Colored People, the sinking of the *S.S. General Slocum* in New York City's East River, the Entente Cordiale between Britain and France, and Austria's annexation of the Turkish provinces of Bosnia-Herzegovina. By the turn of the century, New York State led the nation in manufacturing and the standard of living of its workers was improving. Between 1900 and 1910, its population expanded from 7.3 million to 9.1 million. As the population grew and diversified, so did the circulation of newspapers.

Events that received extensive coverage in New York newspapers between 1911 and 1922 included World War I and the U.S. entry into the hostilities, the Easter Rebellion in Ireland, the Russian Revolutions of 1917, and other changes in Europe brought by the War. Such domestic and local events as the ratification of Constitutional amendments legislating women's suffrage, and prohibition, the Triangle Shirtwaist fire, and the loss of prominent New Yorkers on the Titanic received extensive coverage in the New York press.

Access to New York State Newspaper Collections

Access to New York State newspapers has been significantly enhanced by the NEH-sponsored U.S. Newspaper Project, which, among other objectives, preserves historical newspapers on microfilm. Between 1988 and 1993, The New York Public Library preserved the content of

nearly 500 U.S. newspaper titles with NEH funding. The Library has also contributed to the New York Newspaper Project administered by the New York State Library. The New York Newspaper Project began in 1987 to inventory, catalog, and preserve New York newspaper collections throughout the state. Each title is cataloged to CONSER standards, descriptive information is added to the OCLC Union Catalog, and the holdings are entered into the USNP Union List. The project is also filming newspapers that meet criteria for research value, physical condition, and length and completeness of holdings. The project's website (www.nysl.nysed.gov/nysnp) lists the State newspapers that are currently available on microfilm, as well as holdings information. (Recently, the website was updated and master negative information appears to be missing at this time.)

According to the website, the project has "inventoried 1,759 newspaper repositories, cataloging 22,728 distinct newspaper titles, with 10,503 of those published in New York State, and created 65,903 local holdings records with issue specific data." In addition, working with partner institutions such as The New York Public Library, the project has "microfilmed over 4.37 million pages of New York State history in newspapers amounting to 6,905 rolls of film," supplementing the holdings that exist in archives throughout the state.

In 2005, The New York Public Library was one of only five institutions selected to partner with the National Endowment for the Humanities and the Library of Congress to digitize newspapers from microfilm to create a national online archive of historical newspapers. Since its project began, the Library has digitized nearly 64,000 pages of its goal of 100,253 pages from the 1900-1910 issues of *The Sun* and the *Evening World*. Each newspaper page is scanned in several image and text formats, including Optical Character Recognition (OCR). OCR and metadata created in the project will enable searches of the text and page images. The Library of Congress has developed a beta version of the *American Chronicle* archive, which includes test pages from The New York Public Library's project, as well as contributions from other partners. *American Chronicle* is slated to launch in January 2007.

Many of the other newspapers that are accessible digitally on a large scale are titles that are still commercially available today, such as the *Buffalo News, Daily News, New York Post, New York Times, Newsday, Post-Standard,* and *Wall Street Journal.* Only *The New York Times* and *Wall Street Journal* offer their complete backfiles digitally. *America's Historical Newspapers,* a Newsbank subscription database, has thus far focused on digitizing New York State newspapers from the 18th to early 19th centuries. At the time of this writing, 159 New York titles are included, representing 28 cities. Several specialized historical databases feature a limited range of full-text newspaper articles, such as *The Civil War: A Newspaper Perspective,* which holds selected Civil War related articles from the *New York Herald* between 1860 and 1865. In addition, Brooklyn Public Library digitized the microfilm of the *Brooklyn Daily Eagle,* Brooklyn's newspaper of record. The *Brooklyn Daily Eagle* was published from 1841 to 1955, and again from 1960 to 1963. Brooklyn Daily Eagle Online contains approximately 147,000 newspaper pages from the period 1841-1902 and is provided as a free resource from the Brooklyn Public Library (www.brooklynpubliclibrary.org/eagle). All of the other titles cited above are available digitally for fees.

Methodology and Standards

Selection of Titles

The New York Public Library's holdings of State newspapers, in print and on microform, are extensive, with many titles that were collected solely by the Library or by only a few other libraries. While newspaper collections are dispersed throughout the Library's research divisions, a significant portion is located in the Humanities and Social Sciences Library (HSSL), one of the four Research Libraries in The New York Public Library system. The Humanities and Social Sciences Library's newspaper collection includes more than 350,000 reels of microfilm, tens of thousands of bound volumes, and thousands of current titles from all over the world. HSSL maintains comprehensive collections of most general New York City newspapers in English and other languages. In addition, HSSL provides access to representative U.S. newspapers in many languages from major metropolitan areas, to international newspapers with at least one title from every country when possible, and to other selected newspapers. The Library also subscribes to many electronic resources that index newspapers, often including full-text articles.

When it applied for the initial testbed NDNP grant, the Library identified a number of New York State historical newspapers that met the criteria established by the NEH for coverage, ownership of master negatives, and chronological span. As the center of newspaper production, New York had a large number of prominent newspapers. The titles selected for review were published in English, provided state coverage, and had wide regional influence. Excluded were titles that were purely local, were partisan, or are currently published, as were titles for which the Library did not hold the master negative. The Library also looked at the length of the run to ensure that the titles selected covered much if not all of the time frame for NDNP project. The completeness of the title run held by the Library was also a factor in the selection. A review of the titles under consideration found that the Library's film holdings are more complete than those of other archives; in many instances, the Library was the source of the microfilm that resided at other institutions.

Based on these criteria, the Library selected *The Sun* and the *Evening World* for digitization in the preliminary phase of the project, and proposes to continue with these two titles for the next phase. In selecting the titles, the Library took into consideration not only the criteria cited above, but also the recommendations of the proposal reviewers and its Humanities and Social Sciences Library curatorial staff. Master negatives for a significant portion of the title runs are available at the Library (see Appendix A for details), positioning the Library to fulfill the 100,000-page production goal and to continue to provide content in future phases.

The Sun, launched in 1833, was the first successful penny paper. A daily newspaper, it emphasized human-interest reporting and lively, sensationalist writing, which resulted in high circulation, particularly with the working classes. "Yes, Virginia, there is a Santa Claus," perhaps its most famous editorial, was published in 1897 in response to a letter from a young girl. The Library is the only institution in the state that holds the master negatives for *The Sun*. The Library has the majority of the film for the run between 1833 and 1950, including all the film for 1880 to 1910. The average size of the daily in 1880 was 4 pages. By 1885, it had grown to 8 pages. By 1910, the daily had expanded to 14-16 pages. The Sunday edition of The Sun was 8

pages in 1880 and 12 pages in 1885, growing steadily to 24 pages in 1890 and 30 pages by 1899. By 1910, some issues were up to 40 pages. The 1900-1910 issues are the focus of the Library's current grant-funded NDNP project. The second phase of the grant project will target the 69 reels of master negative microfilm for the period 1880-1899, which contain an estimated 71,000 pages. There is no known paper copy of the full run of *The Sun;* the Library of Congress holds in Remote Storage bound paper issues for September-December 1883.

The *Evening World* began in 1887, an offshoot of *The World*, a morning paper purchased by Joseph Pulitzer in 1883. *The World* appealed to the working classes with its features and human interest reporting, and by 1887, was the most profitable newspaper in New York, with a circulation of 100,000. In March of that year, Charles A. Dana of *The Sun* introduced the *Evening Sun* with the aim of attracting readers of *The World*. By October, Pulitzer had retaliated by introducing the *Evening World*, known within the *World* offices as "Junior." Over time, it became more popular than its older sibling. The Library is the only institution in the state that holds the master negatives for the *Evening World*. The *Evening World* began as a 4-page paper but by 1893, had doubled in size to 8 pages. By 1900, it was 10 pages, doubling in size to 20 pages by 1910. The 1900-1910 issues are being digitized under the Library's current NDNP grant. In the second phase of the project, the Library will digitize issues from the period 1887-1899, of which there are 55 reels of master negatives, containing an estimated 32,500 pages. The Library has not been able to locate any paper copies of the *Evening World*.

More detailed historical overviews for both titles are available as attachments to this application.

Microfilm Quality

Microfilming production varied significantly before best practices were established and widely adopted. The quality of the film held at the Library depends on when the filming was carried out and by whom. With its in-house lab, the Library has been able to consistently apply the prevailing best practices to the microfilm it produces. A sampling of the microfilm for the titles *The Sun* and the *Evening World* showed that the quality characteristics of the microfilm, produced in-house in the 1950s, are consistent with those of film produced in the U.S. Newspaper Project, except for the number of pages in a frame and the reduction ratio. Some of the reels were filmed one page to a frame, and others were filmed two pages to a frame. The reduction ratio for some of the film was found to be slightly above 20x. To ensure accurate scanning, manual review of the reels, both by the Library and by the digitization vendor, will be necessary.

A sampling of the microfilm for the expanded 1880-1910 timeframe exhibited the same characteristics as the film in the Library's current NDNP project. Assessment of the film through a microfilm reader showed that film clarity is very good and that pages are highly readable. However, when film was initially sent to the digitization vendor for scanning in the Library's current project, a portion of it was returned as unscannable. When the reels were inspected under a high-powered microscope, it was concluded that images of text were not as legible on a microscope or finely tuned lens as they were on a microfilm reader. This was due to the resolution and slight variations in density, or the lightness or darkness of the image, that exist for a portion of the materials. To make the materials more scannable, the density was corrected by

the Library's preservation microfilming lab during the microfilm duplication process. The degrees of skew were acceptable, with the skew corrected by the digitization vendor to facilitate the OCR process.

Preservation Reformatting Program

A large number of newspapers have been preserved on microfilm by the Library's own in-house Goldsmith Preservation Laboratory, which can be considered the premier in-house lab in the United States dealing exclusively with preservation microfilm. The onsite microfilming lab, a microfilming preparation section, and a special projects unit comprise The New York Public Library's 15-staff member Preservation Reformatting Program, who work together to prepare materials for reformatting, microfilm the materials, and inspect the film. Staff also contributes to the administration of the Library's vast microfilm collection, one of the largest in the world.

The microfilming preparation section works with the curators and subject specialists in each division of The Research Libraries on their largest preservation task: the microfilming of the intellectual content of the numerous brittle volumes in the general stack collection that are of special research value. Preparation for filming includes collation, searching, queuing, target preparation, recordkeeping, and quality control.

The Goldsmith Preservation Laboratory produces microfilm for the Library's brittle collections, including preservation projects funded by the National Endowment for the Humanities; fills public orders for materials on microfilm; and reformats such high profile collections as "The Nabakov Manuscript and Archive" and the "Western Union Telegraph Company Records." Its microfilming, processing, and quality control equipment permit the Library to maintain national technical standards for all operations and to respond in a timely fashion to the many national and international requests from scholars for copies of the titles in the master microforms collection.

Master negatives produced in the Library's filming projects are stored at NYPL's off-site storage facility, ReCAP. Environmental conditions are controlled at 50°F and 35 percent RH with gaseous and particulate filtration systems. Service copies are stored in the curatorial division or central stacks of the Humanities and Social Sciences Library. The stack storage areas of the Library are monitored on a regular basis by the Facilities Maintenance Office and by the staff of the Goldsmith Preservation Division.

Digital Library Program

The New York Public Library has been at the forefront of research centers in using digital technology to make library collections available to a global audience. It is also a founding member of the Digital Library Federation, collaborating with leading academic and national libraries to explore and implement emerging digital technology. The Library's website (www.nypl.org) provides access to its catalog and electronic databases, together with a growing body of digital reproductions from its collections. Inaugurated in 1994, the NYPL website was used by 19.5 million visitors last year. Visitors to the digital library collections usually account for 16 percent of all traffic to the site, or one out of six visitors.

The Library's commitment to digital library initiatives was formalized in 2000 with the inception of its Digital Library Program (DLP). The 12-person DLP staff produces and manages the content of the Library's digital collections, such as searchable archival finding aids, full-test documents, digital surrogates of materials, and guides to images and "born digital" materials. The Library's collections of multimedia digital holdings include manuscripts, archives, prints, photographs, maps, original art, illustrated books, printed rarities, sound files, and moving images from the four NYPL Research Libraries. In addition, the DLP also sets benchmarks and correlates practices institution-wide for capture, description, storage, retrieval, and delivery of digital materials. One of the leading programs in the field, it participates in ongoing development of best practices in the digital environment.

Digital conversion is performed in the fully equipped DLP Digital Imaging Unit, which photographs and scans objects, creates derivative files, and conducts quality control on the digital products. The Digital Imaging Unit also produces technical metadata, including TIFF tags, capture information, file formats, and identifiers, and manages digital assets through file naming, file transfer, and storage. Access to the images is made possible by the work of Metadata staff, which produces descriptive data records for objects, following national standards for authority versions of names, subjects, and formats. The Metadata staff also creates identifiers, maintains relationships among digital objects and between the objects and their collection-level and item-level metadata records, and links the objects to their catalog records.

The DLP Technical staff manages the content to be displayed on the website by developing online presentations and maintaining text-encoding standards for content. The Technical Encoding staff oversees the development and maintenance of digital library resources such as full text-searchable electronic transcriptions of primary materials, reference databases and indexes, image-based HTML "page turners," and archives of "born digital" resources. Staff also coordinates the use of Encoded Archival Description (EAD) at the Library by developing and maintaining local application guidelines, converting existing archival guides to EAD format, and migrating SGML EAD files to Extensible Markup Language (XML). Furthermore, staff monitors the development and evolution of technical and digital library standards such as XML, Metadata Encoding and Transmission Standard (METS), Text Encoding Initiative (TEI), and open source systems and software.

The Library has undertaken numerous large digital conversion projects, including *In Motion: The African American Experience*, a three-year project funded by the Institute of Museum and Library Services. *In Motion* (www.inmotionaame.org) is a digital archive of primary source materials documenting 400 years of African American migrations to, within, and out of the United States. It features 16,700 pages of text and 8,300 illustrations drawn from books, narratives, manuscripts, articles, essays, and images from the Library's renowned Schomburg Center for Research in Black Culture. These items are presented in 13 exhibits, each a "stream of migration," in which the interpretation of the materials is enhanced by scholarly essays. *In Motion* was selected as one of the 2005-2006 "Best of the Humanities on the Web" honorees by NEH's EDSITEment website.

The *NYPL Digital Gallery* (http://digitalgallery.nypl.org/nypldigital) is a database of 480,000 digital images of artwork, maps, photographs, prints, manuscripts, illustrated books, and printed

ephemera. The materials represent highlights from the Library's vast art, humanities, social sciences, applied sciences, and performing arts collections. Examples of this content include artwork such as Goya's *Disasters of War*; panoramic cityscapes of New York City's Fifth Avenue; classic illustrated zoologies and botanies such as *Pomona Britannica*; historical and cultural images—a vast category represented by such near-contemporaneous works as George Catlin's North American Indian Portfolio, Felice Beato's photographs of Japan, and reformer Thomas A. Larcom's portrait collection from Dublin's Mountjoy prison; theatrical documentation including the Theatre Guild's first production of *Porgy* in 1927; decorative arts in fine pochoir prints of the 1920s; and rare illustrated books such as William Blake's hand-printed masterpiece of 1793, *America a Prophecy*. The site is a rich visual resource for scholarly research and artistic inspiration. The *NYPL Digital Gallery* was named the "Best Research Site" by Museums and the Web 2006 at the annual Best of the Web awards.

The Library also contracts with outside vendors to convert materials, as appropriate. However, when rare or fragile materials need to be digitized, the Library's in-house capacity allows it to capture these materials without removing them from the premises. In instances when the Digital Imaging Unit is fully committed to other projects, the Library will request that contractors scan the materials onsite to minimize risk to these materials. Vendors must adhere to the same guidelines and standards used by the Digital Library Program, which were developed by following best practices promulgated by the Digital Library Federation, Research Library Group, and Library of Congress.

Through its many diverse projects, the Digital Library Program has been able to explore a wide range of technical issues. Staff have been able to investigate system design, method of content retrieval and delivery, hardware and software requirements, emerging technologies and standards, storage and preservation, intellectual property issues management, organizational models, staffing requirements, access policies, and user needs. The technical experience and highly skilled staff of the Digital Library Program, as well as the Goldsmith Preservation Laboratory, make the Library the perfect partner for this ambitious National Digital Newspaper Project.

Work Plan

The process of reformatting microfilm masters of newspaper into usable digital files requires a series of steps that will be completed in part by Library staff and in part by an outside vendor, iArchives, the vendor the Library selected for the preliminary phase of the project. The Library plans to continue to outsource the digital conversion of the materials and the metadata creation to iArchives in order to take advantage of the best equipment available and to draw upon the vendor's growing expertise with the NDNP project requirements, derived from its digitization contracts with the Library and several other NDNP partner institutions.

The Library's staff will perform microfilm review, duplication, and preparation; reel metadata preparation; and quality control in the Library's in-house preservation microfilming lab. Manual review of the microfilm will be necessary so that staff may flag potential problems that will affect scanning. Proper preparation of reel metadata will facilitate the reformatting work by

enabling the clear identification and linking of the digital output with the required electronic records. Quality control is required to vet the various digital outputs and reconcile them with the technical and bibliographic metadata. To facilitate file management and to replicate the validation process used by the Library of Congress, The New York Public Library has created an Oracle database and Fedora repository to accept the data and process the digital files created in this project. The Library will review and validate all files for accuracy before they are sent to the Library of Congress.

Overview of Work Process

Part One—Library staff
1. Award contract to vendor
2. Retrieve microfilm for the two titles
3. Extract bibliographic data from the Online Public Access Catalog
4. Perform technical evaluation of and prep film
5. Assign LCCN (LC Control Number) for bibliographic titles
6. Create reel metadata
7. Create print master for vendor and the Library of Congress
8. Create bibliographic record in HADES (the Library's digital bibliographic database)
9. Ship microfilm print master and reel metadata to vendor

Part Two—Vendor
1. Perform digital reformatting of microfilm in accordance with Library of Congress specifications
2. Create accompanying metadata in accordance with specifications
3. Package digital assets into METS ((Metadata Encoding and Transmission Standard) Submission Information Packet
4. Validate digital files and wrappers to the technical specifications
5. Write content to the Library-provided data drives
6. Ship digital content to the Library
7. Retain print master until validation and ingestion are completed by the Library of Congress

Part Three—Library staff
1. Receive data drives from vendor
2. Upload data drives to working storage
3. Run Library of Congress validator on uploaded batch, as per specifications, and verify that the transfer has not corrupted the shipment
4. Verify content headers, as per specifications
5. Begin process of Library ingest into Fedora repository
6. Verify coherence of shipment
7. Validate content through manual quality control (visual review of 1 out of every 100 pages)
8. Assign unique and persistent identifiers for all digital files
9. Parse files into appropriate folders for processing into archive
10. Accession files into archive
11. Verify files in archive with digital signatures from vendor documentation
12. Create Fedora objects

13. Approve batch
14. Ship approved batch to the Library of Congress for validation and ingest
15. Any errors or mistakes require communication with vendor and the process is repeated until the batch is approved

Part Four—Library of Congress
1. Validate partners' content
2. Ingest partners' content
3. Approve content
4. Notify partners of approval

Part Five—Library staff
1. Repeat process with all batches until complete

Scanning Specifications

In carrying out this project, the Library will adhere to the 2007-2009 technical guidelines set for the National Digital Newspaper Program by the Library of Congress. The newspaper pages will be scanned from second-generation duplicate silver negative film. The contractor will produce 8-bit grayscale images at 400 dpi relative to the original newspaper pages. A minimum of 100,000 image files will be delivered in uncompressed TIFF 6.0 format. In addition, page images will be converted to searchable Portable Document Format (PDF). The PDF Image with Hidden Text will be grayscale and downsampled to 150 dpi and medium JPEG quality setting. Lastly, the Library will provide JPEG2000 compressed image files for each page, with 8:1 compresson, enabling a low-bandwidth presentation of the image.

Microfilming practices have varied through the years, and sampling has shown that, from reel to reel, the number of newspaper pages captured in a frame varied. In instances when a frame contains more than one page, each page will be scanned to separate files. Furthermore, film images with a skew of greater than 3 degrees will be de-skewed by the digitization vendor to facilitate accurate OCR production.

Scanning resolution targets provided by the Library of Congress will be delivered along with the microfilm targets and page images.

Optical Character Recognition

The vendor will also perform Optical Character Recognition on the newspaper pages, conforming to the ALTO XML schema, per the project's technical guidelines. The OCR text will be delivered in UTF-8 character set, and graphic elements on the page will be omitted. The OCR text files will include bounding-box coordinate data at the word level. Bounding-box data associates words to a position on the image and makes possible full-text searching.

A minimum of 100,000 OCR text files will be produced, corresponding to each page image. The text will be segmented by column, reflecting the actual contents of the page, rather than by article. As per the technical guidelines, the OCR will be delivered uncorrected. To ensure that the

OCR output corresponds to the page images in dimension, orientation, and skew, the Library will not make corrections to the images once OCR has been performed.

Metadata

The Library will follow the metadata specifications determined by the Library of Congress. A full bibliographic record for each title is available in the CONSER/OCLC catalog. This catalog record will be the basis for the structural metadata provided for each title. Structural metadata (e.g., page number, date, edition, sequence information) will be provided for each page, issue, and title. Technical metadata, such as the file format, the pixel array, the targets, and the device, will be provided for each page scan, as well as for each microfilm reel used for the digitization (describing the quality characteristics of the film used). The metadata will also reflect the file associations that relate each item.

Technical metadata for the film used for the scans is captured by Preservation Reformatting staff with oversight from Digital Library Program staff. Additional technical metadata will be entered by the vendor at the time of digital capture. The initial metadata is transmitted in the form of a spreadsheet, which is used as a carrier-sheet for the microfilm that goes to the vendor, and also transferred to the Digital Library Program, where it is loaded into the data structure in an Oracle database. When files are returned from the vendor, the issue and page metadata is parsed into both the Fedora repository and the Oracle database.

Methodology

The Library's methodology is informed by its experience in the preliminary phase of the NDNP project. One of the expressed goals of the pilot phase of the project is to develop best practices for the digital capture and processing of materials from microfilm. The Library learned that the reformatting of these materials requires not only collaboration with the digitization vendor but also extensive preparation, tracking, and quality control on the part of the Library itself; and consequently, it has built internal processes and developed software to accommodate these requirements.

The program technical lead, the Library of Congress, provided all participating partners with high-level detailed technical specification and a suite of software tools to facilitate the quality control and verification required by the specifications. It was clear from the start, however, that the requirements of the National Digital Newspaper Project are far more complex than for simple digital reformatting. The reel metadata requirements alone are more extensive than is typical in digital projects where the original format is a bibliographic particular. Furthermore, the range of tolerances for density, reduction ratio, resolution, and other microfilm characteristics were developed in light of post-RLG guidelines and may or may not pertain to the microfilming practices of the 1950-1970s, when the project's newspaper titles were filmed. Consequently, much more detailed examination of the materials was required, and only hands-on evaluation would allow the Library to meet the program and vendor guidelines.

The architecture developed by the Library of Congress demanded that deposits into the digital repository be made in specific digital packets modeled after the Open Archival Information

System (OAIS) reference model and expressed as Metadata Encoding and Transmission Standard (METS) documents. (The METS protocol provides the XML structure that relates and links all the digital elements, allowing for the tracking and migration of digital materials through generations and physical relocations.) Each METS document, containing structural, bibliographic and technical metadata about the microfilm artifacts and the digital artifacts, and the associated deliverables are to be submitted to the Library of Congress *in toto* to be processed as a self-contained unit.

Given these requirements, the Library made the decision to issue a Request for Proposals and outsource the digital reformatting of the microfilm material. The Library determined that its staff resources were best spent in the evaluation of the original content, the management of the high volume of digital files and METS documents flowing into and out of the Library, and the quality assurance of its deposits to the LC repository.

The Library also made a change in workflow process, which was originally modeled on the digital production workflow in place at the Library for the past five years. The initial workplan called for two quality control positions—one for metadata and one for the digital files—and a production manager to handle the distribution of digital assets between the vendor and LC. However, even though the Library had a great deal of experience with mass digitization, mass processing was very different and required some changes to the workflow. The Library of Congress also partnered with the Harvard JHOVE project to customize the JHOVE toolkit for the NDNP project's purposes—enabling the validation of large numbers of digital files from microfilm. The changes to the Library's workflow accommodated the shift from manual handling of single files to automated handling of multiple files, as well as the project's need to track multiple threads of data coming in and out of the Library.[1] The sheer volume of material involved required the Library to build infrastructure to manage the file movements; the validation was so complex that it was more cost-effective and efficacious for machines to do the bulk of the internal coherence checks while visual checks were done manually by staff. The Library shifted its grant resources to accommodate these new processes.

In order to determine whether the vendor-produced content was suitable to be forwarded to the Library of Congress, the Library needed to replicate the validation process employed by the vendor since the purpose of the validation is to guarantee integrity, and the integrity of any file system can be easily compromised by transfer protocols. In the long run, this decision to validate the vendor output was not insignificant since the vendor output was a complex set of interrelated files. However, this afforded the Library the opportunity to test one of the other stated goals of the NDNP project—to provide a testbed for digital preservation and the life-cycle management of digital objects in a repository.

Life-cycle management necessarily involves checks and re-checks at every stage of the life of a digital object. The Library was able to mimic these stages by using the vendor's NDNP-specific output as the testbed. Since the actual deliverables of the NDNP project are processes and practices as well as digitally reformatted newspaper content, this was for the benefit of the project and the long-term benefit of digital preservation. If the Library could manage the

[1] See the workflow included as an attachment

complexities of file handling and management, it would demonstrate that the Library of Congress' model was not a "one-off" achievement and that a digital preservation practice could be developed by LC and implemented by other institutions under their own policy environments.

The initial development stage was to create a series of parsers that could handle the vendor output and authenticate the verification as published. This guaranteed that content sent on to LC was known to The New York Public Library and truly was "content as advertised." In order to accomplish this, the Library created Oracle database tables to accept the data produced by the vendor. The Library engaged the services of a JAVA programmer to create parsers to deconstruct the METS documents provided by the vendor and to populate the Oracle tables. Existing digital archive processes were modified by the Library's Information Technology Group to receive the NDNP files as legal file types to be included in the physical storage areas.

The following were programmatically validated and reviewed before content was sent to the LC repository.

1. Everything referenced in the governing METS document was first checked for coherence. *Is everything (every digital object, text, and raster) that is referenced in the document contained in a referenced subfolder?* If the supplied METS document is not coherent, the parser will fail, the process will stop, and LC will reject the deposit.

2. All digital files were computationally processed through the JHOVE software. Files that passed the validation process were embedded with proper technical headers and this information was encoded into a separate XML document that was also included in the METS governing document.

   In other words, if a file is referenced in a METS document as a TIFF file with a *.TIF extension, the ingest process checks that it is a valid TIFF file and outputs an additional file with a digital signature, a unique numeric value, that is attached to the file and is used as a checkdigit to guarantee that the file has not degraded or changed during transport. This "second pass," where all digital objects are validated against their own self-documentation, is also a coherence test and any level of anomaly prevents deposit into the LC repository.

3. Reconciliation of the documentation against the actual bibliographic content of the master digital file, and the master digital file against its corollaries. *If the METS governing document includes a bibliographic description of* The Sun *for January 2, 1902, and the structural map says it is in subfolder A-9, its corresponding JPEG2000 file is in subfolder B-9, its ALTO OCR document is in subfolder D-9, and the Reel Number is referenced in the top governing folder of the shipment batch, do all these items match (in content)?*

4. Bibliographic integrity. *Is the pagination correct? Is the item correctly described? Is the structural map correct? Is the file legible?*

These steps were taken for several reasons: the specification of the METS submission and the JHOVE software checks changed during the first quarter of the project and items that would have passed the "manual" review would be rejected by LC if an automated process had not been put in place; the need to track and identify specific files could not be accommodated without internal automated processes; and the file naming schema of the METS documents allows for internal coherence but once files are orphaned from their transmission package, the names are not unique, which disallows re-starts when processes are interrupted and renders these "orphans" virtually unmanageable unless they are accessioned into a repository.

Consequently, in addition to creating NYPL-specific software tools for this project, the Digital Library Program implemented a Fedora repository to house its NDNP output and complete the life cycle, from creation to ingestion to migration, in its own digital repository. Somewhat different in architecture from the Library of Congress, the NYPL dataflow[2] stands as a proof of concept for the NDNP program and for the accessioning of digital content in a variety of automated processes smartly strung together to guarantee the provenance of digital content.

The Library has gained tremendous new expertise by engaging in the NDNP project. The project has allowed the Library the opportunity to create a preservation testbed using a critical amount of content. It has enabled the Library to manage one of the most complex bibliographic structures, newspapers, and to demonstrate that the OAIS reference model has practical application. And it has allowed the Library to evaluate the requirements of mass processing and of tracking, handling, and migrating data, and to create internal systems to manage these needs.

Schedule

Following is a schedule that reflects the sequence of project activities and the staff responsible for completing them. Please note that the schedule reflects the activities that will be undertaken by project staff, including the shipment of batches to the digitization vendor, but the details of the digitization are included earlier in this Work Plan section. Additional information about the digitization, which will adhere to the technical guidelines provided by the Library of Congress, is included in the vendor's quote, provided as an attachment to this application.

<u>First Quarter (July 2007-September 2007)</u>

Project Co-directors (Director, Digital Library Program / Chief, Preservation Division)

a) Attend program meeting/workshop at the Library of Congress.
b) Convene kickoff meeting for project staff to review workplan and schedule monthly meetings.
c) Contract with digitization vendor.
d) Hire Library Technical Assistant to carry out microfilm inspection, digital file quality control, and data transfer procedures.
e) Establish project guidelines.
f) Coordinate activities across Library divisions.
g) Liaise with project partner, Library of Congress, National Endowment for the

[2] See attached Dataflow diagram.

Humanities, and other award recipients to administer the project.

Preservation Reformatting Head and Supervising Camera Operator

a) Recall microfilm materials and perform assessment of condition, noting areas of concern, including missing issues.
b) Duplicate microfilm reels per grant project requirements

Special Projects Associate, Preservation Reformatting

a) Assess available bibliographic description.
b) Enter microfilm reel metadata on carrier-sheet.

Metadata Coordinator

a) Extract catalog records in XML to a staging area where they are parsed and reorganized for loading into Oracle.
b) Develop digital file naming conventions. Assign unique identifiers to item-level objects.
a) Review carrier-sheets and extract reel metadata to upload to Oracle server.

Senior Web Developer

a) Update Intranet site to document and share project information among Library staff

XML-Java Developer

a) Monitor evolving Library of Congress technical specifications. Modify validation processes locally to comply with any specification revisions affecting incoming vendor-produced batches.

<u>Second Quarter (October 2007-December 2007)</u>

Head, Digital Imaging Unit

a) Train Library Technical Assistant on the quality control benchmarks for materials digitized from microfilm.
b) Upgrade and/or modify validation software on network storage.

Library Technical Assistant

a) Assist in performing condition assessment on microfilm.
b) Under supervision of Head of Digital Imaging Unit, ship first batch of collection for testing and scanning (up to 10,000 pages).

Supervising Camera Operator

a) Duplicate microfilm reels per grant project requirements.

Metadata Coordinator

c) Extract catalog records in XML to a staging area where they are parsed and reorganized for loading into Oracle.
d) Develop digital file naming conventions. Assign unique identifiers to item-level objects.
b) Review carrier-sheets and extract reel metadata to upload to Oracle server.

Senior Web Developer

a) Develop functional requirements and test use cases for an interface that will allow non-technical staff to view and query digital newspaper objects stored in the Fedora repository.
b) Update Intranet site to document and share project information among Library staff.

XML-Java Developer

a) Write Java GUI (graphical user interface) to Fedora software, or adapt GUI from available open-source projects.

Chief, The Irma and Paul Milstein Division of United States History, Local History And Genealogy

a) Write decade essays for project, per grant requirements

Project Co-directors (Director, Digital Library Program / Chief, Preservation Division)

a) Monitor production progress.
b) Complete quarterly report.

Third Quarter (January 2008-March 2008)

XML-Java Developer

a) Oversee ingest of vendor-produced metadata and content, as well as parsing to appropriate Fedora and Oracle storage servers.

Metadata Coordinator

a) Review carrier-sheets and extract reel metadata to upload to Oracle.
b) Review vendor-produced metadata that has been parsed and loaded into Oracle and Fedora repository.

Head, Digital Imaging Unit

a) Implement validation software packages.
b) Oversee the flow of digital objects into archive.
c) Monitor quality control procedures and policies.

Library Technical Assistant

Under supervision of Head of Digital Imaging Unit:

a) Receive first batch of digital files (TIFF, JPEG2000, PDF, OCR) and metadata from vendor.
b) Send second batch of microfilm to vendor (up to 15,000 pages).
c) Conduct visual quality control on sample of records for TIFF, JPEG2000, PDF, and OCR files.
d) Run derivative processing for internal review.

Senior Web Developer

a) Liaise with repository developer and the Library's Oracle consultant for validation of vendor-produced content.

Project Co-directors (Director, Digital Library Program / Chief, Preservation Division)

a) Monitor production progress.
b) Complete quarterly report.

Fourth Quarter (April 2008-June 2008)

Head, Digital Imaging Unit

a) Implement validation software packages.
b) Oversee the flow of digital objects into archive.

Senior Web Developer

a) Liaise with repository developer and the Library's Oracle consultant for continued validation of vendor-produced content.

Library Technical Assistant

Under supervision of Head of Digital Imaging Unit:

a) Receive second batch of digital files and metadata from vendor.
b) Send remainder of microfilm to vendor.
c) Conduct visual quality control on sample of records for TIFF, JPEG2000, PDF, and OCR files.
d) Run derivative processing for internal review.

XML-Java Developer

a) Oversee ingest of vendor-produced metadata and content, as well as parsing to appropriate Oracle and Fedora storage servers.

Metadata Coordinator

a) Review carrier-sheets and extract reel metadata to upload to Oracle.
b) Review vendor-produced metadata that has been parsed and loaded into Oracle and Fedora.

Project Co-directors (Director, Digital Library Program / Chief, Preservation Division)

a) Monitor production progress.
b) Complete quarterly report.
c) Disseminate and promote project results.

Second Year (July 2008-June 2009)

Project Co-directors

a) Accept final digital packages from vendor.
b) Attend program meetings/workshops at the Library of Congress.
a) Monitor production progress.

b) Submit grant reports as required.
c) Ship microfilm print masters to Library of Congress.
d) Disseminate and promote project results.

All Project Staff

a) Continue organization and quality control processes with remainder of digital files.

Staffing

Please see the attached resumes and position descriptions for details.

Barbara Taranto, Project Co-director / Director, Digital Library Program. With co-director, will have oversight of all aspects of the project, including financial management and reporting. Will facilitate the work of project staff across Library divisions. Will liaise with the project partner, funding agencies, and contractors. Will develop procedural guidelines for project, particularly in the area of digitization. Will assist the Operations Manager, as needed. Ms. Taranto is Director of the NYPL Digital Library Program, and is responsible for the administration of digital materials, including collection development, data preservation, and application software development to ensure the quality and interoperability of the Library's electronic collections. Ms. Taranto will contribute 10% of her time to this project.

Evelyn Frangakis, Project Co-director / Aaron and Clara Greenhut Rabinowitz Chief Librarian for Preservation. With the co-director, will have oversight of all aspects of project (see other Project Co-director responsibilities described above). Will develop procedural guidelines for project, focusing on the area of preservation microfilm. The Aaron and Clara Greenhut Rabinowitz Chief Librarian for Preservation, Ms. Frangakis, is responsible for general administration, planning, and supervision of all aspects of the Library's preservation program, including close monitoring of preservation treatment and production. Ms. Frangakis will contribute 10% of her time to this project.

Saskia Scheffer, Head, Digital Imaging Unit. Will be responsible for overseeing the work of the Library Technical Assistant. Responsible for the shares on the Digital Library Program's networked storage, the flow of digital objects into the archive, and the implementation and oversight of the NDNP batch validator and visual validation interface. Will supervise the quality control of digital assets and processes. Ms. Scheffer will be assigned to this project at 15% for the duration of the grant period.

Janet Murray, Metadata Coordinator. Will review metadata for accuracy and advise other project staff on technical issues related to the creation, storage, and retrieval of digital items. As Metadata Coordinator, Ms. Murray has oversight of all metadata creation in the Library's digital conversion projects. Ms. Murray has 20 years of archival experience. She will contribute 25% of her time to this project.

Joseph Dalton, Senior Web Developer. Will implement web-based interfaces for tracking completed reels and contributing data for the new metadata fields for this project that leverage

the Library's Fedora repository, Metadata Creators Interface, and other project-management tools such as BaseCamp. Mr. Dalton is a Digital Library Program Senior Web Developer who develops software tools for digital library collection development and user access. He provides support for user interfaces and database applications, advises technical developers on industry standards, and is responsible for coordinating the work of programmers, software engineers, and consultants on the Library's public digital interfaces. Mr. Dalton will contribute 10% of his time for the duration of the project.

Richard Lane, Head of Preservation Reformatting. Will assess the condition of the microfilm materials, and note any areas of concern, including missing issues. Will also create the necessary microfilm copies for this project. Mr. Lane is responsible for developing, coordinating, and managing the Library's preservation reformatting programs and staff. Mr. Lane will contribute 5% of his time to this project, primarily in a coordination and oversight capacity.

Library Technical Assistant I. Will perform physical inspection of the microfilm prior to digitization, and perform quality control on the digital images and the metadata returned from the vendor. Will receive data drives from digitization vendor and will upload data drives onto specified shares on the network attached storage. Will monitor the progress of the data transfers and report any anomalies to the Head of the Digital Imaging Unit. Will confirm completion of data transfer. Will ship validated drives to the Library of Congress, making sure all appropriate carrier sheets and documentation are included. Will track the physical hard drive devices to and from the vendor and the Library of Congress. Will keep accurate records of data transfer times and dates and participate in monthly project meetings. This position will be created for this project.

XML-Java Developer. The Library will contract with a XML-Java Developer to write the Java graphical user interface (GUI) to the Fedora software, or adapt the GUI from available open-source projects. In addition, the XML-Java Developer will oversee the ingest of metadata and content, as well as parsing to the appropriate Oracle and Fedora storage servers. The XML-Java Developer will also monitor the Library of Congress technical specifications and modify validation processes to comply with any revisions that affect the batches.

Conclusion

The Library is well-prepared to continue its digital conversion project, and it has the infrastructure, resources, and expertise necessary to sustain this effort in future phases. The Library's deep collections, and long-standing involvement in the NEH/U.S. Newspaper Project, ensure that it will have ample materials upon which to draw. Highly skilled Library staff has led and participated in numerous digital conversion projects, including the preliminary phase of the NDNP project. The staff's experience in managing the administrative and technical aspects of those projects will guide and inform the ongoing effort to digitize materials in microfilm format. The Library looks forward to contributing its rich historical holdings and its technical expertise to the next phase of this national initiative.

The New York Public Library
National Digital Newspaper Project, 1880-1910
July 1, 2007 - June 30, 2009

Year I - Fiscal Year 2008

1. SALARIES AND WAGES		%	Annualized Salary	NEH Cost	Cost Share	Total
Position Title	Incumbent					
Chief Librarian for Preservation	E. Frangakis	10%	[redacted]		[redacted]	[redacted]
Director, Digital Library Program	B. Taranto	10%				
Head, Digital Imaging Unit	S. Scheffer	15%				
Metadata Coordinator	J. Murray	25%				
Senior Web Developer	J. Dalton	10%				
Chief, US History Division	R. Carr	5%				
Head, Preservation Reformatting	R. Lane	5%				
Supervising Camera Operator	T. Jean-Louis	10%				
Special Projects Associate	L. Moureaux	30%				
Library Technical Assistant I	New position	50%				
Subtotal Salaries and Wages				**70,111**	**29,149**	**99,260**
2. FRINGE BENEFITS						
Full-time @ 38.5%				26,993	11,222	38,215
Subtotal Fringe Benefits				**26,993**	**11,222**	**38,215**
3. CONSULTANT FEES						
Java Developer				7,500		7,500
Subtotal Consultant Fees				**$7,500**	**$0**	**$7,500**
4. TRAVEL						
To Washington DC for annual meetings	2 people/1 trip @ $350 pp/per trip			$700		700
Subtotal Travel				**$700**	**$0**	**$700**
5. SUPPLIES AND MATERIALS						
Supplies for production of print masters				$4,500		4,500
Subtotal Supplies and Materials				**$4,500**	**$0**	**$4,500**
6. SERVICES						
Digitization	$0.85/page * 50,000 pages			$42,500		42,500
Subtotal Services				**$42,500**	**$0**	**$42,500**
7. OTHER COSTS						
Shipping to/from vendor and LC				$500		500
Subtotal Other Costs				**$500**	**$0**	**$500**
Total Direct Costs				**152,803**	**40,372**	**193,175**
Indirect Costs @ 7.5% (*)				10,095	3,028	13,123
TOTAL COSTS - Year I				**$162,899**	**$43,399**	**$206,298**

(*) IDC only applied to first $25,000 of digitization contract

The New York Public Library
National Digital Newspaper Project, 1880-1910
July 1, 2007 - June 30, 2009

Year II - Fiscal Year 2009

1. SALARIES AND WAGES		%	Annualized Salary	NEH Cost	Cost Share	Total
Position Title:	Incumbent:					
Chief Librarian for Preservation	E. Frangakis	10%	[redacted]		[redacted]	[redacted]
Director, Digital Library Program	B. Taranto	10%				
Head, Digital Imaging Unit	S. Scheffer	15%				
Metadata Coordinator	J. Murray	25%				
Senior Web Developer	J. Dalton	10%				
Chief, US History Division	R. Carr	0%				
Head, Preservation Reformatting	R. Lane	5%				
Supervising Camera Operator	T. Jean-Louis	10%				
Special Projects Associate	L. Moureaux	0%				
Library Technical Assistant I	New position	50%				
Subtotal Salaries and Wages				**58,626**	**25,158**	**83,783**
2. FRINGE BENEFITS						
Full-time at 38.5%				22,571	9,686	32,257
Subtotal Fringe Benefits				**22,571**	**9,686**	**32,257**
3. CONSULTANT FEES						
Java Developer				7,500		7,500
Subtotal Consultant Fees				**7,500**	**0**	**7,500**
4. TRAVEL						
To Washington DC for annual meetings	2 people/1 trip @ $350 pp/per trip			700	0	700
Subtotal Travel				**700**	**0**	**700**
5. SUPPLIES AND MATERIALS						
Subtotal Supplies and Materials				**0**	**0**	**0**
6. SERVICES						
Digitization	$0.85/page * 50,000 pages			42,500	0	42,500
Subtotal Services				**42,500**	**0**	**42,500**
7. OTHER COSTS						
Shipping costs (to/from vendor & Library of Congress)				500		500
Subtotal Other Costs				**500**	**0**	**500**
Total Direct Costs				**132,397**	**34,843**	**167,240**
Indirect Costs @ 7.5% (*)				6,690	2,613	9,303
TOTAL COSTS- YEAR II				**139,086**	**37,456**	**176,543**

(*) IDC only applied to first $25,000 of digitization contract

The New York Public Library
National Digital Newspaper Project, 1880-1910
July 1, 2007 - June 30, 2009

Summary

1. SALARIES AND WAGES		%	NEH Cost	Cost Share	Total
Position Title	Incumbent:				
Chief Librarian for Preservation	E. Frangakis		0		
Director, Digital Library Program	B. Taranto		0		
Head, Digital Imaging Unit	S. Scheffer			0	
Metadata Coordinator	J. Murray			0	
Senior Web Developer	J. Dalton			0	
Chief, US History Division	R. Carr		0		
Head, Preservation Reformatting	R. Lane			0	
Supervising Camera Operator	T. Jean-Louis			0	
Special Projects Associate	L. Moureaux			0	
Library Technical Assistant I	New position			0	
Subtotal Salaries and Wages			**128,736**	**54,307**	**183,043**
2. FRINGE BENEFITS					
Full-time @ 38.5%			49,564	20,908	70,472
Subtotal Fringe Benefits			**49,564**	**20,908**	**70,472**
3. CONSULTANT FEES					
Java Developer			15,000		15,000
Subtotal Consultant Fees			**15,000**	**0**	**15,000**
4. TRAVEL					
To Washington DC for annual meetings			1,400	0	1,400
Subtotal Travel			**1,400**	**0**	**1,400**
5. SUPPLIES AND MATERIALS					
Supplies for production of print masters			4,500	0	4,500
Subtotal Supplies and Materials			**4,500**	**0**	**4,500**
6. SERVICES					
Digitization	$0.85/page * 50,000 pages		85,000		85,000
Subtotal Services			**85,000**	**0**	**85,000**
7. OTHER COSTS					
Shipping to/from vendor and LC			1,000		1,000
Subtotal Other Costs			**1,000**	**0**	**1,000**
Total Direct Costs			**285,200**	**75,215**	**360,415**
Indirect Costs @ 7.5% (*)			16,785	5,641	22,426
TOTAL PROJECT COSTS			**301,985**	**80,856**	**382,841**

(*) IDC only applied to first $25,000 of digitization contract