NATIONAL ENDOWMENT
FOR THE HUMANITIES

SAMPLE APPLICATION NARRATIVE

Preservation and Access Humanities Collections and Reference Resources
Institution: 1 HZ < RUN 3 XEOIF / IEUDV

Project Significance

The paper and digital New York City map collections of NYPL's Lionel Pincus and Princess Firyal Map Division serve the research needs of a wide audience, including geographers, cartographers, archaeologists, historians, ecologists, architects, urban planners, sociologists, genealogists, collectors, attorneys, novelists, artists, dramaturges, and screenwriters, among many others. These groups view NYPL's map collection through their unique perspectives and nuanced understanding of landscape, be it topographical, societal, or historical.

Over the past few years, a flourishing of both high-end Geographic Information Systems (GIS) tools and vernacular applications (like Google Earth and Google Maps) have pushed geospatial methods more broadly to the fore of humanities research. This "geospatial" turn offers the potential both for researchers to ask new questions and present the humanities in new modes to both scholarly and popular audiences. A number of very recent initiatives indicate a coalescing interest in geography as method, from this summer's Andrew W. Mellon Foundation-funded Scholarly Communication Institute (with the theme "geospatial tools in the humanities") to the recently awarded NEH Advanced Topics in the Humanities Institute "Enabling Geospatial Scholarship."

In order to truly embrace the potential of geospatial research, a scholar needs access to both the the map and the data that the map represents, which requires steps beyond simple digital imaging. For example, once imaged, each map in a given atlas must be "georectified" to a base map (so that a location on one map can be correlated with the same location represented on another map). Then, the non-geographic information (borders, page numbers, etc.) must be cropped out, and the individual pages aligned with each other to form a single holistic layer. Next, the building footprints, parks, and other shapes represented on each map must be traced as vectors, so that they can be machine-readable and thus useful for computational analysis and visualization (in a process similar to what Optical Character Recognition (OCR) offers for text materials). Finally, any additional information, such as building height or place names, must be entered and correlated with the relevant vector shapes.

Once maps are more comprehensively digitized in this vein, they open up new avenues for humanities scholarship on a scale previously impossible. For example, researchers could:

- Conduct a study of the historical evolution of the built environment with an eye towards 19th century architectural styles, building materials, and resource depletion in the region;
- Build three-dimensional models of New York City for literary, performing, and visual artistic interpretation of the historical landscape;
- Reconstruct the city, using historical building heights as part of a larger study of the particular history and effects of the urban landscape on the human condition and psychological state;
- Analyze the material types of residential structures to infer social class patterns on a microcosmic level, cross-referencing to contemporaneous census records;
- Locate and survey new potential archeological sites;
- Conduct highly granular studies of the effects of new rapid transportation lines on the built environment; or

- Unlock genealogical information by using historical ward and census boundary data combined with old street name and addressing schema.

The data to support such work, however, can be incredibly time-intensive to develop. While computational techniques like edge-detection offer some promise for homogenous groups of maps, for the most part this is still human work, requiring labor far beyond the capabilities of any institutional staff (which is why most collections of georectified maps number in the hundreds, not tens of thousands). The Library will address this limitation through a unique approach, described below.

There are over nine thousand images of maps in the NYPL *Digital Gallery*, the Library's repository of 700,000-plus digitized images, but virtually none had gone through a more substantial process of georectification and data digitization until 2007 and 2008. At that time, the Library began working with students from Pace University, who initially georectified more than 800 digitized maps on a volunteer basis. Building on that initial success, NYPL invested in the development of a web-based platform for scaling up the sophisticated digitization of maps by spreading the work across more volunteers, in essence "crowdsourcing" the georectification and data extraction to any and all participants. In the Spring 2009 semester, Pace University students georectified another 800 maps. This model, much like the "citizen science" movement in astronomy, ornithology, and other observational sciences, opens up the possibility of creating geospatial data on a previously impossible scale, as well as encouraging greater participation by the public in building resources for research and scholarship. NYPL will implement this model in proposed project activities.

The work supported by *NYC Historical GIS* will both significantly extend and accelerate the process by which geospatial researchers gain access to Library's heavily used collection, both as images of historical maps and as GIS data traced from historical maps, and also establish a new model (including both infrastructure and process) that will be useful for other such collections.

Project History, Scope, and Duration

The Lionel Pincus and Princess Firyal Map Division

The Lionel Pincus and Princess Firyal Map Division (http://www.nypl.org/research/chss/map/map.html) is one of the world's premier map collections in terms of its size, scope, unique holdings and diversity of use. Established in 1898, holdings include over 430,000 sheet maps and 16,000 atlases and books representing every corner of the earth from global to, occasionally, the architectural scale.

A core holding of the Map Division is the antiquarian atlas and map collection, which includes many important early Dutch, English and French imprints dating to the 16th century and is supported by strong corpus of secondary resources held in the Library's various Divisions, including the General Reference Division, Manuscripts and Archives Division, and Photography Collection, for its use and study. This collection was formed when the Astor and Lenox Libraries merged to form The New York Public Library and was later augmented by significant gift collections including the Ford Collection, John Levine Bequest, Emmet Collection,* and Lawrence H. Slaughter Collection.

Another key holding is topographic map sets, some from the founders' libraries, including: César-François Cassini's *Carte de France Levée par ordre du Roy*, a topographic map set covering pre-Revolution France in 175 sheets; the *Ordnance Survey of Ireland's Townland survey of the counties of Ireland*, a 35-volume detailed survey of early 19th century Ireland; and the Austro-Hungarian Empire's *Spezialkarte der österreichisch-ungarischen Monarchie*, published in the late 19th and early 20th centuries and covering the empire in 773 sheets. The bulk of its topographic maps, however, came to the NYPL during the 20th century as part of the Federal Library Depository Program with significant map coverages added through aggressive collection development policies aimed at deepening strengths for the entire world.

Maps and atlases detailing cities throughout the world represent another core strength of the Map Division, with the historical New York City map holdings among the deepest and most heavily used anywhere. With more than 2,000 sheet maps and 13,000 atlas map sheets documenting the city and its five boroughs before 1922 (often to the building level) , this collection has been a critical support to many research projects throughout the division's history and will be where work in this proposal is focused.

Researchers and scholars as well as organizations use the historic New York City map holdings for a variety of research purposes, and are guided to these resources by NYPL curatorial and subject specialist staff. For example, urban archeologist Joan H. Geismar, Ph.D., regularly uses the Library's historic New York City map collection to uncover clues to a past landscape that no longer appears on maps. Dr. Geismar's work has proven instrumental in uncovering and telling of many of New York's early stories. This information often finds its way into exhaustively researched site reports, such as an archeological assessment of Washington Square Park, which helped inform the recent reconstruction of the Park.

Staff of *The Mannahatta Project* used a 1902 folio of New York City geology maps from the Map Division as a basemap layer to determine the historical surface geology of Manhattan Island. The research informed a complex ecological model of Manhattan Island as it would have appeared before the arrival of Europeans. When related to the project's other historical data such as elevation, slope, hydrography and other data types, the Mannahatta team was able to reasonably hypothesize what soils, flora, fauna and human landscape elements might have occupied every corner of the island.

Similarly, visiting Fellows of NYPL's Dorothy and Lewis B. Cullman Center for Scholars and Writers have frequently consulted the Map Division, many enhancing their work through accessing detailed maps of New York City. Columbia University Professor Farah Jasmine Griffin conducted research on Black women artists in New York using the Map Division's property maps of Harlem from the 1920s to 1940s to learn more about area cultural and social locales such as theaters, hotels, row houses, and hospitals. For author Nelson Smith's research for a biography of a tenement house, curators retrieved property atlases of the Lower East Side. Similarly, for his pending work on popular crime, Smith consulted maps detailing the location of the butcher shop of the murderous William Poole (also known as 'Bill the Butcher').

Map Digitization

For over 100 years, the Map Division has provided traditional library services to the public, and since the 1960s, it has operated out of its magnificent reading room located in the Stephen A Schwarzman Building in midtown Manhattan. Over the last decade, however, the staff of the Map Division have been pushing content beyond the walls of the Library by working in concert with NYPL's Digital Imaging Unit to publish close to 10,000 map images to the web.

The Division's digital work started with the gift of the Lawrence H. Slaughter Collection and its mandate to provide web access to this collection of English maps and atlases of the Mid Atlantic Seaboard of Colonial North America. In 2001, NYPL received an NEH grant to conserve, catalog, and digitize over 1,000 of these maps, which are now part of the larger NYPL *Digital Gallery*, but are thematically preserved in the website *American Shores Maps of the Middle Atlantic Region to 1850* (http://www.nypl.org/research/midatlantic).

The Division subsequently participated in the Institute of Museum and Library Services-funded (IMLS) project "Building A Globally Distributed Historical Sheet Map Set" with the University of Connecticut (as lead institution) and the American Geographical Society (http://imlsmap.lib.uconn.edu/about.html). The project sought to digitize, georectify and serve via web map server, a mosaicked topographic map set of the Austro-Hungarian Empire published between 1877 and 1917, which brought together paper maps held in the three separate institutional repositories. The project modelled a scalable infrastructure for sharing scanned, geo-referenced images of sheet maps in a metadata-driven environment. For this project, NYPL digitized, georectified, clipped and presented 773 maps with attendant metadata records.

Over the last several years, the metadata creation and digitization have been integrated into the Divisional workflow with Division staff entering metadata records directly into the NYPL's Oracle database. As a result, the digital map collection has grown to close to 10,000 images. The entire body of digitized maps can be viewed through the Library's *Digital Gallery* at http://digitalgallery.nypl.org/nypldigital/ dgdivisionbrowseresult.cfm?trg=1&div_id=hm.

Preliminary Work: Extending the Utility of Digitized Map Content

As part of a strategic, long-term plan to extend the reach and utility of its digitized map content, in summer 2008 NYPL contracted with EntropyFree LLC, a geospatial software design firm, to build an open source web map server as well as georectification and vector tracing tools. The tools allow staff and external users to transform the digital map collections: first, into comprehensive digital historical atlases (as map mosaic images) and then into digital map data. These lightweight mapping tools are easy to use with very little training, unlike a sophisticated GIS application, which typically require extensive training on steep learning curves.
Once maps have been digitized within the established NYPL workflow, the georectification tool enables users to georectify or electronically "stretch" historic maps over a contemporary base map (see illustration below). The georectified maps are then available from the web map server on the NYPL website and exported as Web Map Service (WMS) layers accessible using desktop GIS software or as layers in Google Earth. (To view an example of a georectified map, go to: http://dev.maps.nypl.org/warper/ mapscans/10196 and click on the "Preview Rectified Map" tab.)

Example of a georectified historical map stretched over a contemporary map

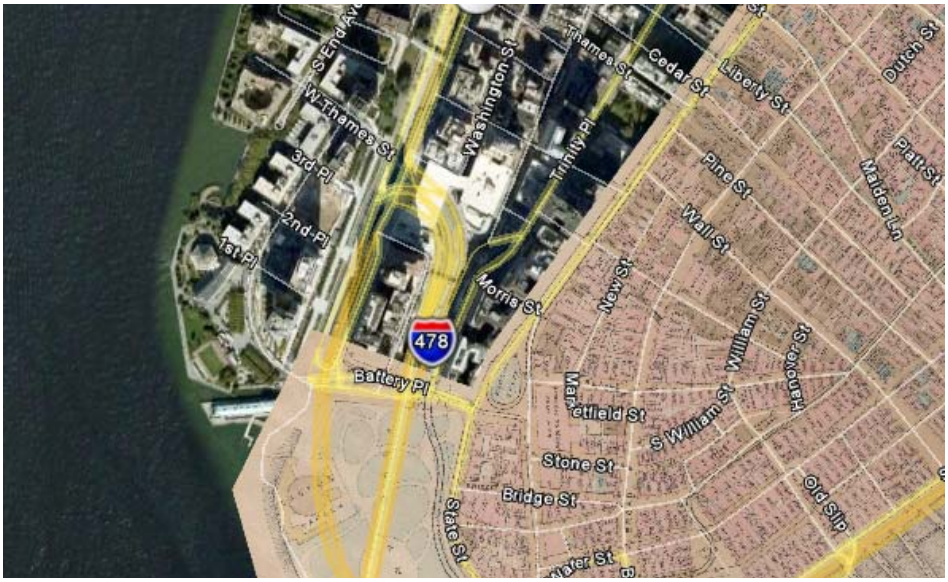

Plate 1 from Elisha Robinson's *Atlas of the City of New York,* 1885 overlayed in Google Earth

Completion of the initial web map server in late 2008 dovetailed with a pilot crowdsourcing project, as described above, by which students georectified maps using commercially available GIS software. The *NYC Historical GIS* Project Director oversaw 20 interns from a Pace University community service/computer science class who, by May 2009, had received training and georectified close to 800 of NYPL's historic New York City property maps. These maps are the first set of maps that project staff will be repurposing as GIS data layers using the vector tracing tool described in detail below.

The result of these crowdsourcing activities is available in a Google Maps-like interface with historical map images "stretched" and draped over a contemporary map. The interface has functionality users have come to expect such as pan-and-zoom as well as a transparency and a soon to be enabled time control, allowing the user to fade the historical image in and out of vision and scroll through maps of the same location through time. Please see Appendix A-1 for a screenshot of the homepage of the historical maps web platform and A-2 for a screenshot of a rectified map, laid out over the actual geography it depicts.

The georectification tool also enables users to crop out non-map margins from individual maps that are part of a larger map set. Maps can then be pieced together into seamless mosaics in which multiple map sheets are digitally "stitched" together into a single plane. An example of where this is quite useful is with William Perris' *Maps of the City of New York*, 1857-62, a seven-volume atlas covering Manhattan Island from lower Manhattan to 72nd Street in 122 separate maps, each covering an area of several blocks. A user of the digital mosaic copy of this atlas, instead of flipping back and forth through from map to index pages in the paper atlas, would browse, via the NYPL web map server, a seamless version, moving from page to adjacent page, without knowing or needing to know which page was to the north or south. Please see Appendix A-3 for a visual of William Perris' *Maps of the City of New York*, 1857-62, each page rectified, aligned and laid out in Google Earth.

In the third phase of the Library's ongoing work, to debut in fall 2009, EntropyFree developed a tool for users to trace geospatial information datasets from georectified maps. Shapes traced from a map, a building for example, become containers of information within a geographic dataset and are annotated with data about the source map as derived from the map's metadata record, and any other information depicted on the source map. For example, the traced map could include the name of the school, the building material, the number of stories and/or the building address. Additionally, in this process, staff categorize the type of feature, in this example, "schools," which is then added to a larger dataset of schools derived from all different mosaic map sets. These larger feature sets would represent components of a historical geographic data model which in turn will become the framework for providing geographic access to other non map collections.

Working within the Library's existing digitization workflow and utilizing the tools described above, the Library seeks funding to continue building on this initial work through a new project focusing on a significant portion of NYPL's historical New York City map collections. Through the three phase process of digitization, georectification, and tracing outlined below, the Library will build a significant new type of reference work critical to the study of humanities in this densely layered, ever-changing City.

Grant Project Description: *NYC Historical GIS*

Collection Digitization

Project staff have selected materials to be digitized as part of this project according to both the needs of researchers, who often seek historical New York City maps, as well as from NYPL's collections strategy, which aims to develop comprehensive digital resources for New York City. With this dual mandate, the NYPL has to date digitized 77 fire insurance atlases, totaling more than 3,100 map sheets, which detail the New York City environment to the building level dating to 1852. These previously digitized fire insurance atlases and approximately 9,500 maps to be digitized as part of the project itself will form the foundation of proposed project activities.

Fire insurance maps are among the most heavily used in the Division, and are notable for their great detail. The fire insurance mapping industry was born in the aftermath of the great New York City conflagration of 1835, which destroyed 20 million dollars worth of property. The fire bankrupted many small insurance companies and brought about substantial changes in the way insured property was identified and described. Beginning with the earliest productions, the format of the maps has stayed relatively constant, measuring slightly smaller than 2 x 3 feet, bound or in single sheets. NYPL's fire insurance map collection is unique in that it contains manuscript notations from previous owners, including possibly a real estate office. These notations appear as slips of paper pasted to the maps or as direct handwritten revisions; often they indicate multiple house numbers, showing when and where renumbering took place. Local historians, architects, archaeologists and urban planners are all avid users of these historical maps of New York City at the block and lot level. They use the maps to research the history of particular buildings and neighborhoods, for purposes of social history and design, as well as for ecological and engineering reasons. Sometimes the maps speak directly about the histories of the companies or homes documented, and sometimes they facilitate research more obliquely.

For practical reasons, developers and architects are eager to learn why there is water in their new building site; using these maps they can determine that an underground stream is the problem, or perhaps an old swamp or wetland area, such as that which Canal Street, once a canal, was built to drain. Cemeteries long ago built over are locatable; and brownstone owners can determine whether or not wooden porches did grace the back of the row houses on their block.

Scholars use the same information for different purposes. For example, a researcher reading and studying the diaries of Herman Melville's family, held in NYPL's Manuscripts and Archives Division, found a passage referring to a deck where, in his old age, penniless and depressed, Melville often sat, rocking in his rocking chair, overlooking a pianola factory. A sheet map from William Perris' 1862 *Maps of the City of New York* showed not only the house, but also the porch and the pianola factory. Another researcher, an author, writing a book on the case of the Lindburgh kidnapping, used fire insurance atlases of Staten Island to locate the garage where the kidnappers stashed their loot. The maps themselves, while originally intended for use by the fire insurance industry, now provide a dramatic backdrop to, and are full of clues about, the City's rich and detailed history. The same history unfolds and is written, photographed, and recorded in countless other formats throughout libraries and personal collections everywhere.

As noted above, the first stage of *NYC Historical GIS* project staff will digitize approximately 8,000 atlas map plates from the Map Division's fire insurance atlas collection described in greater detail below. This includes titles published by Matthew Dripps, D.A. Sanborn, E. Belcher Hyde, William Perris, Elisha Robinson, G.W. Bromley, and F.W. Beers. Additionally, funding will support the digitization of close to 1,500 sheet maps covering each of the City's five boroughs. Digitization will take place over the course of the entire grant period, with 3,000 maps digitized during each of the first two years and 3,500 in the third year. All maps included in this project were published between 1851 and 1922 and are in the public domain. The majority of these materials are considered in good condition. The project's plan of work takes into account minor instances of brittle or torn pages, and staff have budgeted for minor conservation work.

Atlas Map Sheets

The 153 atlases to be digitized during the grant period represent New York City during a period of phenomenal and unprecedented urban growth. The core of the 64 titles for which the Library has page counts (5,109 atlas map sheets) are Sanborn Fire Insurance maps, known for their extraordinary level of detail. In addition to showing all of the built structures in any given location, these maps outline, among many other things, the locations of overhanging building cornices and the exact locations of sidewalks, elevators, skylights, dumbwaiters and laundry chutes. This series of atlases, published between 1887 and 1918, covers Manhattan, Brooklyn, and part of the Bronx and came to the NYPL from a project at the Library of Congress to distribute duplicate atlas copies originally deposited for copyright. An additional 95 titles containing approximately 2,850 atlas map sheets include atlases published by Bromley, Robinson, Hyde, Dripps, Perris, Perris & Browne, Crofton, and Spielman and Brush. This collection, published between 1860 and 1922, covers the Bronx, Brooklyn, Manhattan, and Queens. The number of atlas map sheets is based on an estimated average of 30 sheets per volume for these particular publishers.

Loose Sheet Maps
Along with the atlas map sheets, the NYPL will digitize 1,500 loose sheet maps of the five boroughs drafted by myriad publishers. These maps, primarily of smaller scale (less detailed, wider coverage area) than the fire insurance maps, typically cover an entire borough on one sheet. Often times, however, they predate the fire insurance maps and therefore represent a critical, albeit less detailed step back in the historical landscape record. An E. Belcher Hyde map of Queens, for example, published in 1907, outlines all of the farm and property boundaries with owner's names and acreages for the then mostly agrarian borough. Additionally, the maps often include thematic information such as early transport lines (horse drawn rail) that can be best understood at the borough wide scale.

Georectification

Selection of Materials to Georectify
For the second stage of *NYC Historical GIS*, project staff will georectify, mosaic, and publish Web Map Service (WMS) layers for 2,000 atlas map sheets, covering all five boroughs. During the first year the project team will georectify a set of 502 previously digitized maps of Queens and Staten Island covering the period from 1898 and 1917, during a time when the population of Queens tripled and that of Staten Island nearly doubled.

During the second and third years of the grant period, an additional 1,500 maps, digitized in the first and second year will be georectfied and published. The NYPL will focus on the Sanborn Fire Insurance maps mentioned above, selecting a subset of maps dating from between 1900 and 1913 and covering the Bronx, Brooklyn, and Manhattan. During that historical period, the population of the Bronx more than doubled, while Brooklyn and Manhattan gained close to one-half million residents each. Manhattan, by 1910 had 2.3 million residents (600,000 more than today), with new immigrants crammed into tenements concentrated on the Lower East Side.

With the immense flow of humanity into the newly consolidated City of New York, a tremendous effort to remake the city was undertaken that included massive demolition and construction projects to house new residents. The maps digitized and georectified during for this grant represent a lost snapshot, an earlier version of the City constantly remaking itself.

Tracing

Selection of Materials to Trace
In the project's third stage, staff will trace geospatial information, specifically building structures and property boundaries from digitized, georectified maps representing each of the five boroughs of New York City.

The initial set of 518 maps to be traced during the first year of the grant period encompass the City's five boroughs during several distinct time frames. The following comparison of data to be traced from two atlases of Manhattan from two nearly adjacent time frames illustrates the rapidity and dramatic nature of changes to the early City. William Perris' *Maps of the City of New York*, the first two series of fire insurance atlases for the borough of Manhattan, were published between 1852 and 1854 and again between 1857 and 1862. These maps represent a period of rapid growth and development in the lower half of the island. Between 1850 and 1870,

the borough population nearly doubled, reaching close to one million residents. This growth is clearly etched on the pages of these two publications with entire blocks of brownstone rowhouses quickly erected on the neatly laid out street grid. An earlier, quieter landscape is evidenced by the occasional wooden framed farmhouse sitting out of alignment to the same grid. Or, in the east side Kips Bay neighborhood, entire sections of old houses aligned to a street grid laid out in the Colonial era awaiting demolition and reordering as the development of the new city marches uptown. These maps are the most detailed evidence in the Library's collections of the built environment as recorded in the mid-19th century. A scholar can easily visualize these changes, without the hours and weeks of research it would take to identify the paper maps (of varying pagination from several publishers), image and crop them so they appear on the same page, and layer them alongside and on top of one another.

The digitized maps selected to be traced during the second and third years of the grant (see the Plan of Work section) offer up a comparable wealth of data from a variety of time frames representing periods before, during and after massive change. For example, comparing William Perris' 1855 Maps of the City of Brooklyn, with D.A. Sanborn's Insurance Maps of the City of Brooklyn, (1904-1908), offers the picture of two radically different places that occupy the same geographic space, the former a maritme city, locus of many early homesteads, prominent family domiciles, and leading businesses important for the study of the history of New York City to a huge, modern metropolis, second only in size to Manhattan with trolleys, road and streetcar networks uniting far neighborhoods carved from what was once the most productive agricultural county in the nation.

Similar Collections

When geography is understood as an organizing framework for the history, society, and culture of New York City, one begins to see threads of geographic interconnection woven through the myriad digitized materials within NYPL collections, including but not limited to real estate photographs and historical apartment floorplans from the Irma and Paul Milstein Division of United States History, Local History, and Genealogy; Berenice Abbott's WPA photographic series *Changing New York* and Lewis Hine's documentary photographs of Ellis Island and the construction of the Empire State Building, located in the Miriam and Ira D. Wallach Division of Art, Prints and Photographs; and the *Buttolph Menu Collection*, containing more than 9,000 historical menus from mostly New York City restaurants housed in NYPL's Rare Books Division. While it is beyond the scope of proposed grant activities, all of these materials contain geospatial and temporal data that will one day be correlated to the data structure created by *NYC Historical GIS*. Together, these materials will provide extremely nuanced and detailed insight into how New York City was shaped over two centuries.

Similarly, the maps processed as part of this project relate in a tangible way to collections throughout other New York City agencies and organizations, collections that are firmly rooted in concepts of "place" and "time". These materials include, but are not limited to: land conveyance records in each of the borough registrar offices dating to the early Dutch period; atlases and maps of New York City in both the Brooklyn and New York Historical Society libraries; detailed topographic surveys in the offices of each of borough's Topographic Bureaus; nearly 900,000 digitized historical photographs of buildings and municipal records of Brooklyn's original six Dutch townships at the NYC Municipal Archives; 1 million hand-drawn certificate of occupancy

apartment and house floorplans in NYC's Housing and Preservation Department; collections and reports from the Landmarks Preservation Commission; the archives of the Brooklyn Navy Yard Corporation; and historical epidemiological records housed at the Department of Health and Human Services. Existing digital projects such as NYC Open Accessible Space Information System (OASIS), a partnership that maps GIS data on green spaces in New York City, also provide valuable opportunities for partnership and resource-sharing. NYPL is committed to and actively cultivating relationships with all of these institutions with an eye towards future collaborative efforts (See Appendix B: "Letters of Support"). The outcomes of this grant will provide the geospatial data and technical infrastructure to enable such partnerships to flourish.

Methodology and Standards

To improve intellectual access to the project materials and ensure their long-term survival, NYPL is committed to applying the best current professional practices and standards for archive, digitization, and library practice in all its work. The standards and methodologies to be employed in this project cover digitization; descriptive and geospatial metadata creation; preservation; rights clearance; and selection criteria.

Descriptive Image Metadata

The Library's catalog provides the fullest and most structured access to the collections and provides the basis for bibliographic metadata for digitized materials. Once selected for digitization, catalog records are extracted in XML format to a staging area where they are parsed and reorganized for loading into an Oracle database, and ultimately into the Library's Fedora repository alongside binary files. The bibliographic records (once in XML format) are assigned unique identifiers in the database, where each identifier refers to a bibliographic record.

New bibliographic entries or additions will be managed under the direction of the Library's Metadata Coordinator, who will serve as a resource person for Library staff in the proper use of the data entry and retrieval screens. Naming conventions and identifiers will be automatically generated as items are digitized, technical metadata will be added. The Library will also add administrative metadata based on the location and access of the digital library. The METS protocol is used to provide the XML structure to track and migrate digital materials through several generations and physical relocations.

The creation of additional item-level descriptive metadata will be administered through the Library's Metadata Creators Interface, a web-based application that allows for remote data entry and classification using a suite of tools including searching, creating, updating, linking and editing metadata records. The Library's metadata standards are in accord with all Dublin Core Standards.

Geospatial Metadata

Working from the descriptive metadata records generated for the digitization and management of the map images, the Library will create geospatial metadata records for each derivative composite atlas layer (published as WMS) and property/structure layer (published as Web Feature Services) in accordance with the Content Standard for Digital Geospatial Metadata

(CSDGM), Vers. 2, by the FGDC. Additionally, they will follow guidelines outlined by the ISO 19115 "Geographic Information - Metadata"[1] from ISO/TC 211.

NYPL has conducted a due diligence search on the Geospatial One-Stop (GOS) Portal and has determined that needed geospatial-related data for the NYC Historical GIS project does not exist. The Library will use the US Geological Survey's Geospatial Metadata Validation Service to check the metadata against the core elements of the CSDGM.

Digitization

The digitization of images will be managed through the Library's Digital Imaging Unit (DIU). The Unit currently uses a variety of capture devices. Archival images will have a grayscale and color (Kodak Q13 or Q14) bar, and metric or inch ruler placed along the margins of the long dimension, these may be layered to save margin space. Service files will be cropped to extract all extraneous image margins so that no cropping of the physical item shall occur for any files.

Project-supported DIU staff will digitize archival image files at 12 bits per pixel or higher and converted, processed and written to files at 8 bits per pixel depth. The DIU will, at the scanning stage, employ the use of black and white levels, gamma, and look-up tables to provide complete control of the composition, color, and tonal range and values of the image as it is digitized and processed. The service files may be sharpened to a level 4 and scaled during post processing to pre-process them for the later production of derivative files for web presentation. The archival and service images will be written in TIFF format with TIFF version 5.0 or 6.0 Header. The archival files will be retained as the files of record and a unique and persistent identifier will be assigned to the digital object to ensure continued access to the electronic content.

Technical metadata including the file format, the pixel array, the color space, the targets, and the device will be entered at the time of digital capture; these specifications will be stored in an Oracle database in XML format to ensure greatest flexibility of delivery. The compressed web images will be cut by the Library's DIU. Thumbnails will be written in *.GIF and enlargements in *.JPG format, respectively. All web files will be transferred to the Digital Library Program's web server and indexed through an Oracle database. The DIU will be responsible for inspecting the accuracy of file names and directories for all digital images produced under their supervision. Digital image quality will be managed by the DIU. All DIU products are vetted by a quality control process (100 percent examination, file opening and statistical sampling after transfer.

Preservation

The Library has made provisions for indexing and organizing digital files for long-term preservation, accessibility, and usability. The Digital Imaging Unit generates a unique identifier for every file of record, this identifier is maintained in a memory map and is a re-direct to either real storage or a proxy server, depending on user permissions and application. The Digital Imaging Unit will generate an MD5Checksum for every file of record. The numeric identifiers will be statistically sampled before and after migration or re-location and bit checking will be done routinely.

The Library houses its archival TIFF images in a 60 TB extensible Storage Area Network; digital files produced in the Digital Imaging Unit or received from outside vendors are processed through a series of steps. The first step is verification of visual quality, color fidelity, and acuity. Files that pass the visual review are then cropped (color bars and rulers removed) to produce a service tiff, which is then sent for sampling. Four derivative files, one gif, two jpegs, and one SID (wavlet) file are produced from the service file. All derivatives are then visually inspected and arranged into packets of six files, one archival, one service, one gif, two jpegs and one SID, which are deposited in a "hot" folder to be accessioned into the archive. Each file is verified for its format integrity and MD5 Checksums are generated. All files and file dimensions are logged into the FileStore database, which manages the hard storage, and archival files are written to the archive SAN. Derivative files are written to the presentation servers for quick delivery to the web, presentation servers are load balanced, and there are three complete copies of derivative files. Archival storage is backed up incrementally daily and complete backups are run on a regular schedule, and finally, all tapes are taken from the backups are shipped off site to a secure facility.

Additionally, the Library is in the process of implementing a Fedora repository to manage the technical and structural metadata for its digital assets. The Fedora repository will also eventually replace the FileStore applications currently run elsewhere. All Fedora objects that are currently in the repository, 500,000 total file records, conform to the PREMIS specifications for Digital Preservation metadata. The Fedora objects have also been validated, verified, and digitally signed (signature in Fedora) by the JHOVE set of software.

Bit integrity and long term preservation of digital assets is a serious concern for the Library. Currently the Library plans to implement new preservation practices that modify the backup and offsite storage procedures to include regular electronic measuring and evaluation of hard storage. Bibliographic data for digital objects is stored in an Oracle database. Data structures and data values are routinely backed up to tape and sent off site, restores have been exercised and disaster preparedness is in ongoing development.

Selection Criteria and Intellectual Property

The Library will select materials for this proposed project based on a combination of factors, primarily the Library's unique materials, its overarching collection strategies, and the project's primary audience of humanities researchers, teachers, and students. Staff will select materials that are significant for illuminating issues in history, American studies, sociology, psychology, archeology, literature, and the visual and performing arts. Given the Library's vast collection of rare and/or unique maps, and the vast and comprehensive collections in its larger holdings, staff will be able to provide insight into countless subjects, unavailable through any other resource or collection, via this project.

NYPL has determined that all materials digitized for this project are in the public domain. All metadata created (both by NYPL staff and by participating members of the public) will be licensed under an open license still to be determined. The selection of this license will follow current best practices for open metadata projects (Open Street Maps or Wikipedia, for example).

Geospatial Data

Since early 2008, the Library has been developing a web-based platform for the creation of geospatial data using its already-imaged historical maps. The application, currently available at http://dev.maps.nypl.org, takes a high-resolution image of a map from the existing Digital Gallery and models it through a series of workflow steps:

1) Rectification: A streamlined interface enables users to place "pins" on the historical map, and correlate those pins with the relevant locations on a base map. For New York City maps and atlases, the New York City GIS data is used as the base layer (so that all geospatial information created is interoperable with the City's official data). Please see Appendix A-4 for a visual of the rectification interface (historical map on left, base data on right).

2) Cropping & Alignment: Not all of a given map image is necessarily geographic in nature - borders, titles, and other non-cartographic information must be cropped out using a second tool (which allows the user to simply trace the boundary of the cartographic information). As many of the maps being digitized come from atlases and other multi-page formats, this tool also enables staff to align multiple cropped pages into a single master image, comprising potentially hundreds of individual street-level maps.

3) Tracing: In the GIS community, "digitization" refers not to the imaging of maps, but to the tracing of building footprints, city boundaries and other shapes that are represented in maps. Another interface in the Library's web-based platform comprises a simple, point-and-click toolkit for tracing these shapes, thus creating vector data that can be repurposed by researchers. Because the map image has already been rectified to the base map, any such "shape layers" are normalized across the entire archive, and can be mixed and matched depending on the researcher's need.

The final step in the GIS process is the relation of other information, intrinsic or extrinsic to the map, to the points and shapes that have been traced. This process only possible once the points and shapes are molded into a base of structured, time-enabled geographic information. With the base layers created during the project period, it will be possible to construct a map interface capable of searching for both historical maps and non-map digital materials. Using such an interface, a researcher will be able to open a map search window, find the geographic location he or she is researching, and choose layers of placemarkers linked to other digital objects associated with that place, such as photos of buildings, menus from local restaurants, old prints depicting street scenes, texts written about the area, newspaper articles written about a particular street corner. While this particular functionality will not be built during the proposed project period, it will not be possible without the foundational *NYC Historical GIS* data. The work proposed in this project will not only push the NYPL towards that long term goal, but will further extend the use of NYPL's collections in humanities research as well as open countless avenues for collaboration and further study.

Crowdsourcing

The Library will hire a Library Technical Assistant III (LTA), to be supported by the project, who will be responsible for georectifying, tracing, and data scraping the digitized maps. Initially, this staff member will be directly responsible for the tracing and data scraping of

digitized maps; however, through the first year of the project the team will open up the web-based platform and, following established best practices drawn from other contexts, will seek to actively cultivate a broader community of contributors.

Inspiration for this model comes both from the general online space, in which resources like Wikipedia and Open Street Maps have been built by tens of thousands of committed volunteers, as well as the world of citizen science, where datasets of immense value to researchers have been built by birdwatchers counting migrating birds and amateur astronomers scanning the skies. On one hand, the project team will actively work to publicize and recruit amateur researchers and local NYC history enthusiasts through traditional press and the Library's established social media channels; at the same time, the team will experiment with different ways of incentivizing this work, from an outreach event at the Library to public recognition mechanisms like a “top contributors” scoreboard.

If the crowdsource model is successful, the LTA funded by the grant will increasingly shift to a focus on quality control (vetting and monitoring data generated by the community) and community management/outreach, coordinating the production of many times the data the LTA could create on his or her own; if unsuccessful, he or she will still have created several years worth of geospatial data by the end of the grant, enough to be valuable for future researchers.

Plan of Work

ACTIVITIES	YEAR 1 5/1/2010-4/30/2011	YEAR 2 5/1/2011-4/30/2012	YEAR 3 5/1/2012-4/30/2013
Digitizing	2,395 atlas map sheets Dates:1887-1917 Coverage: Brooklyn *Published on Digital Gallery by end of Y1*	2,479 atlas map sheets Dates: 1892-1912 Coverage: Bronx & Manhattan *Published on Digital Gallery by end of Y2* 1,000 atlas map sheets Dates: 1860-1922 Coverage: 5 NYC boroughs *Published on Digital Gallery by end of Y2*	2,000 atlas map sheets Dates: 1860-1922 Coverage:5 NYC boroughs *Published on Digital Gallery by end of Y3* 1,500 Loose sheet maps Dates: 1851-1922 Coverage: 5 NYC boroughs *Published on Digital Gallery by end of Y3*
Georectifying	502 atlas map sheets Dates: 1898-1917 Coverage: Queens & Staten Is. *Published as WMS by end of Y1* * Digitized *before* grant	746 atlas map sheets Dates: 1904-1908 Coverage: Brooklyn *Published as WMS by end of Y2* * Digitized in grant Y1	777 atlas map sheets Dates: 1900-1912 Coverage: Bronx & Manhattan *Published as WMS by end of Y3* * Digitized in grant Y2
Tracing	518 atlas map sheets Dates: 1852-1909 Coverage: All 5 NYC boroughs *Published as vector data by end of Y1* * Digitized *before* grant * Georectified *before* grant	338 atlas map sheets Dates: 1903-1919 Coverage: Queens *Published as vector data by end of Y2* * Digitized *before* grant * Georectified in grant Y1	164 atlas map sheets Dates: 1898-1917 Coverage: Staten Island *Published as vector data by end of Y3* * Georectified in grant Y1
Dissemination	Conferences: WAML/ALA	Conferences: ICHC/RBMS Publish Y1 GeoMetadata 5/2011	Conferences: ESRIUC/Where2.0 Pubish Y2 GeoMetadata 5/2012
Administration	Recruit/Hire LTA III position by 8/2010	LTA III shifts to Quality Assurance role	LTA III contues Quality Assurance role
Training	LTA III training 8/2010-9/2010 Citizen Cartographer Event 10/2010	Ramp Up Crowdsourcing Citizen Cartographer Event 10/2011	Ongoing Crowdsourcing Activities Citizen Cartographer Event 7/2012
Technical	Add web server B to scale up Map Platform to cluster Expand storage by .5TB	Expand storage by .5TB	Refresh 5 year-old web server A with new hardware Expand storage by .5TB

Staffing

NYC Historical GIS Project Director **Matt Knutzen** has served as the assistant Chief Librarian of The Lionel Pincus and Princess Firyal Map Division since 2001. Mr. Knutzen manages daily operations of the Division and assists in the development of the Division's long-term vision of service to researchers and scholars. He regularly provides individual GIS consultation/tutorial to the public and has produced maps for major exhibitions at NYPL, including *Russia Engages the World*, *Cities in the Americas*, *Ehon: The Artist and the Book in Japan*, and *Letters to Sala: A Young Woman's Life in Nazi Labor Camps*.

Mr. Knutzen regularly makes map library and cartography presentations to a variety of public and professional audiences, most recently at the meeting of the Rare Books and Manuscripts Section (RBMS) of the Association of College and Research Libraries (ACRL) in June 2009. He holds a MA in Fine Arts (2000) with an emphasis in abstract cartography from Pratt Institute,

Brooklyn, New York, and a BA in Geography from the University of California at Berkeley. He is an active member of numerous cartographic organizations, including the NACIS (North American Cartographic Information Society), the International Map Trade Association, the Western Association of Map Libraries, and the Rare Books & Manuscripts Section of the Association of College and Research Libraries.

In his role as Project Director, Mr. Knutzen will maintain oversight of the project and coordinate the selection, sequence, and flow of materials being digitized. Once digitized maps are composited, into a Web Map Service (layer) layer, he will disseminate appropriate metadata information. He will supervise two Library Technical Assistant III (LTA III) positions working directly on the project, and oversee the quality of the metadata produced.

Project advisor **Joshua M. Greenberg** is the Library's Director of Digital Strategy and Scholarship and charged with overseeing NYPL's Digital Experience Group, which is staffed with technical practitioners as well as traditionally trained librarians with a variety of experience or expertise in digital media. Dr. Greenberg, who joined the Library in 2007, was formerly the Associate Director, Research Projects, of George Mason University's Center for History and New Media, where along with other work he co-directed the Zotero project. He has broad experience and understanding of both the content and research needs of scholars as well as new technology-based information services and products that can support scholars and other researchers. Dr. Greenberg holds a Ph.D. in Science & Technology Studies from Cornell University (2004). He will identify different avenues to garner support for the crowdsourcing project and help establish and guide its processes and workflow. In addition, Dr. Greenberg will oversee the development and iteration of the map server, making sure that the server is refreshed and built out as needed. He will ensure that the server is fully interoperable with the metadata repository currently in development.

Janet Murray has been the metadata coordinator of the Digital Library Program and the New York Public Library since 2001. She is responsible for metadata structures and data for the Digital Gallery, the National Digital Newspaper Project and other digital initiatives at the Library. Before becoming Metadata coordinator, Janet worked for many years as Visual materials archivist at New York Public Library, the New-York Historical Society, and the Oregon Historical Society. She taught the "Historian and the Visual Record" at NYU's archival program, and has taught courses relating to visual materials cataloging and metadata issues. She earned a BA from Whitman College in Walla Walla, Washington in 1978 and an MA in History and an MLS both from Indiana University in Bloomington, Indiana in 1983. Ms. Murray will monitor the metadata during the digitization, which will require less than one percent of her time.

Saskia Scheffer has served as the Head of Digital Imaging Unit since 2004. She is responsible for file management in a production environment, implementation of required changes to accommodate new projects and an increased workload, imaging material from the Research Libraries to be included in the Digital Gallery, reworking previously produced files to better fit gallery requirements, the production of high quality derivatives from archival files for web delivery, recommendations for upgrades and changes in equipment and procedures, ad hoc requests from curators and divisions for high quality digital files for publication, and training and supervising photographers and production staff. Prior to this position she was the Coordinator Photographic Services and Permissions at the Library from 2000 to 2004, and from 1994 to 2000

she was Manager of A Different Light, a bookstore in New York. Ms. Scheffer has a BA in Photography from Empire State College and a BA in Linguistics from the University of Amsterdam. Ms. Scheffer will coordinate and manage the digitization of the project, which will require approximately ten percent of her time during the three-year project period.

Mitsue Nagase Balan is a Library Technical Assistant IV in the Digital Imaging Unit and is responsible for producing service filers for web-delivery and general quality control. Prior to her work at the Library she served as a photo technician at Taranto labs, a Japanese language teacher at Berlitz, an editor at the Benesse Corporation, and was a freelance photographer. She earned a BA in Business Administration from Konan University in Kobe, Japan in 1989. Ms. Nagase Balan will spend 29 percent of her time on the project during the three-year project period. **Edilberto Punsalan** is a Library Technical Assistant IV in the Digital Imaging Unit and is responsible for both the digital capture and processing of original files. Mr. Punsalan holds a Bachelor of Science from the University of the East (Philippines) and first worked in the DIU as a Digital Technician before his promotion in 1999 to LTA IV/Photographer. Mr. Punsalan will spend 15 percent of his time on the project during the three-year project period. **Krzysztof Jakubek** has a Master of Science in Architecture from Tadeusz Kosciuszko Technical University, Krakow, Poland, and has worked since the 1980s as a professional photographer and photography instructor. He has worked as an LTA IV/Imaging Technician in the DIU since 2004. For project-related activities, he will be responsible for equipment trouble-shooting and upkeep and high-end digital photography; he will spend approximately 27 percent of his time over the three-year period on project-related activities.

A **Library Technical Assistant III** (LTA III) to be hired for 33 months of this project, will spend 100 percent of his or her time on the workflow of georectifying, tracing, and data scraping the digitized maps. Initially, this staff member will be directly responsible for the tracing and data scraping of digitized maps; however, through the first year of the project the team will open up the web-based platform and, following established best practices drawn from other contexts, will seek to actively cultivate a broader community of contributors. (Please see Appendix D: Job Description.)

Dissemination

Successful dissemination of this project will involve several parallel streams; the primary focus will be on enabling access to and cultivating scholarly use of the digitized historical maps and geospatial data that will result from this work, while secondary initiatives will focus on building a public user community around the ongoing work of rectifying and tracing digital maps as well as evangelizing the ‘crowdsourced’ model throughout the library and archival community.

Scholars are still in the early stages of understanding how to apply geospatial methods to humanistic research, and an important part of the work of the Map Division at The New York Public Library over the next few years will be supporting and growing the community of use for historical geospatial data. The Library will actively seek out opportunities to cultivate use of this resource as well as other evolving geospatial datasets through existing systems and directed outreach. On one hand, as WMS layers of mosaic historical atlases are published, the catalog records augmented with online linkages will be resubmitted to OCLC for inclusion in the WorldCat database, effectively opening up the raw data to the broader global library community

through existing discovery systems (for example, when a user who has Google Earth installed on their computer clicks on a catalog link, the appropriate mapset would open in Google Earth laid out over the contemporary satellite imagery).

To develop a base of user for this and similar materials the Library will, through its participation in the upcoming NEH-funded Institute on Enabling Geospatial Scholarship at the University of Virginia and other emerging venues for collaboration, offer this dataset as a pilot collection for emerging practices in geospatially-enabled humanistic scholarship. The library will publicize the availability of the digitized maps and geospatial data via the homepage of the website once the materials are linked to the *Digital Gallery*, regular alerts in the Library's RSS feeds, and written about extensively on the Map Division blog. Finally, the Library will feature the project in its quarterly magazine for donors and through *NOW*, the Library's newest publication designed to keep patrons updated on projects and happenings.

In addition to the digitized maps and geospatial data, the methodology and processes honed over the course of this grant will be of interest and use to the broader Map Library community. Toward that end, the Library will target professional organizations, societies, conferences and listservs that will provide mechanisms for the dissemination of information, methodologies and knowledge gained through the completion of the project. Project team members have well-established relationships and have participated in a leadership capacity with most of these organizations, ranging from map library organizations to institutions promoting the digital humanities.

With over 380 members, American Library Association's Map & Geography Roundtable (MAGERT) is the world's largest map library organization. It provides a forum for people interested or involved in any aspect of map or geography librarianship. Project staff will seek to romote the NCGIS project through publications such as MAGERT's newsletter *base line* and the *Occasional Paper* series. *Occasional Papers* are reference books that contain information about maps that would otherwise not be available.

Another outlet is the Western Association of Map Libraries (WAML) is an independent association of map librarians and other people with an interest in maps and map librarianship. The Association's *Information Bulletin* is issued three times a year and enjoys worldwide readership. Staff may try to place the project in the *Bulletin*. The Northeast Mapping Organization (NEMO) is dedicated to serving as a unifying body for all who use, produce, collect, and market maps and cartographic information in the Northeast United States; increasing communication between all interested in maps; and working with state, regional, and national organizations and government agencies in dissemination of maps and cartographic information. The *NEMO Newsletter* is published quarterly and contains news and activities of NEMO members, as well as useful information on resources and events in the world of map-making, map librarianship, and geography.

In order to achieve wider exposure to *NYC Historical GIS*, project staff will also seek to work with the Electronic Cultural Atlas Initiative (ECAI) and the International Cartographic Association.

Geospatial metadata will be published in established repositories such as the ALA MAGERT, WAML, and the Geospatial One Stop as well as local repositories such as the New York State

Metadata Clearinghouse as well as ECAI MD Clearinghouse. Links to the online map images will be created automatically by NYPL's Integrated Library System and a process by which links to kml files appear in ther catalog records will be established. Once these links are written, the MARC records will be republished on Worldcat and with all hosted links to kml, harvested and indexed by google servers. Because the mosaic atlas layers and the traced data layers will be created on the open layers platform, they will be interoperable with any other implementation. One such potential venue is the aformentioned OASIS, a New York Public Interest Research Group (NYPIRG)-funded New York City mapping platform primarily focused on open space. This focus is shifting however as, just recently, layers of historical GIS layers created by the Manahatta Project team have been seamlessly integrated into their web platform.

Project staff will use all available avenues of contact to disseminate information about the project, and garner support for the crowdsourcing aspect of project activities. The Project Director will host “Citizen Cartography” events during each year of the project, through which he will promote the project to a broad audience of scholars and researchers. The Project Director and LTA IV position will frequently offer training and “train-the-trainer” sessions to acquaint the research community, academic institutions, and geography enthusiasts with the project’s themes and goals.