ABSTRACT
Project Title: Boston Streets: Mapping Directory Data
Submitted by: Tufts University Digital Collections and Archives (DCA)

Responding to requests for visual information about places and for contextual information about photographs is among the most time-consuming reference work done at any repository with visual images. Libraries and archives are constantly seeking new and more efficient ways to supply this information to casual as well as scholarly patrons. Cataloging and making reference copies of visual images has always been a time-consuming process. This problem became acutely obvious when Tufts began working with the Bostonian Society on a project to create web access to the Society's image collection. Digitizing images for reference use only partially solves the problem, because digitized images must also be discoverable through search engines, and this often leads to extensive and cumbersome data organization projects.

People who seek information about a place, especially information about visual images of that place, should be able to retrieve both specific and contextual information relevant to the place in which they are interested. The Boston Streets: Mapping Directory Data project seeks to demonstrate that organizing all types of informational content around geographic location allows us to make connections between a place and contextual information about that place without significantly expanding the process of data organization.

We will demonstrate that it is possible to organize data by spatial location and then link this data to many other pieces of data that share the same location without having to extensively catalog them by subject, name, or other subjective terminology. At the same time, any subjective terms that are added become discoverable though their relationship to the location.

We will digitize eleven Boston city directories between 1865 and 1955 as the foundation for contextual information about the city. The city directories contain richly detailed information about occupations, home ownership, city government, and commerce. Advertisements throughout the directories provide a history of commerce, advertising, and manufacturing.

Building on the GIS (geographic information system) and electronic text creation skills developed in the Edwin Bolles Archive on the History and Topography of London (an IMLS-funded project), we will use this project to extend the capabilities of the Tufts Digital Library GIS interface to include not only gazetteer information as in the London project, but directory information as well.

The data and the access tools built to deliver both the geographic and text content will be compatible with our FEDORA-based repository system, and will be available to any institution wishing to use them.

Combined with the collaboration between Tufts and the Bostonian Society, we see this project enhancing the ability of both institutions to manage access to their visual collections and provide more efficient service to patrons. This project has the added benefit of giving historians and neighborhood groups new and more versatile tools to document and understand the changes in the city and their neighborhoods over time.

NARRATIVE

National Impact
Boston is one of the nation's most historic cities and the subject of some of the earliest examples of urban photography in the United States. Boston is also one of the best-documented cities in the country, with a wealth of information covering its entire history. The Boston Streets project will not only provide access to material hitherto unavailable electronically, but will demonstrate the usefulness of a geographically-based approach to directory information and to collaboratively held collections. The availability of directory data combined with visual data in a geographic interface is a unique combination of resources and will be a valuable tool for researchers, students, and casual historians.

Adaptability
Building on the GIS (geographic information system) and electronic text creation skills developed in the Edwin Bolles Archive on the History and Topography of London (an IMLS-funded project), we will use this project to extend the capabilities of the Tufts Digital Library GIS interface to include not only gazetteer information as in the London project, but directory information as well. City directories contain richly detailed information about occupations, home ownership, city government, and commerce. Advertisements throughout the directories provide a unique history of commerce, advertising, and manufacturing. We have found in our work with the Bolles collection that starting electronic text collections with reference works such as directories, dictionaries, and encyclopedias supports research in the narrative works that are added later.

The data and access tools built to deliver both the geographic and text content will be compatible with our FEDORA-based repository system, and the data, these tools, and documentation about them will be available to any institution wishing to use them.

The collaboration between the Bostonian Society and Tufts University can be a model for bridging the gap between the needs of smaller institutions and their ability to support their projects. Tufts' hosting of the metadata and the GIS interface will allow the Bostonian Society to improve access to its collection material while also allowing Tufts students and researchers anywhere access to material in both collections. This collaboration can serve as a model for a larger-scale collaborative effort among other historical and educational institutions in the Boston area. The development of metadata standards for describing collection contents across institutions, especially between historical societies and research libraries and the potential for creating a common metadata repository with a GIS interface will create the potential for an interoperable database of directory and visual information gathered from the many resources in the city.

The combination of the mapping directories project and the separately funded photo cataloging project enables us to develop and extend the interoperability of data and enhance the value of both projects by offering another means of access for architectural historians, urban historians, urban planners, students, and people interested in

documenting the growth of their neighborhoods. For example, Chinatown Banquet, a non-profit organization dedicated to preserving and exhibiting the history of Boston's Chinatown has recently used images from both collections in exhibits about the history of its neighborhood. Their research necessitated visits to multiple institutions, and would have been well served by the availability of the images in a geographically-based union catalog.

This project will create a model for a larger consortial effort that could move beyond Tufts, and we are actively seeking input from others in the city doing work with images and maps (See Appendix B: Letters of Support). Tufts is also an active member of the Boston Library Consortium (BLC) a cooperative association of sixteen academic and research libraries that includes the Boston Public Library, Northeastern University, Boston College, and Boston University. The BLC has just launched a new digitization projects task force to explore collaborative possibilities among the institutions. The Boston Streets project will be one of the first items discussed by the BLC task force.

Design
The project will have two elements:

1. Digitize 11 Boston City Directories from 1865 - 1955 for years ending in "5" in the collection of Tufts University and the Bostonian Society and make them available as image front, full text searchable documents available from the Tufts Digital Library. We choose the years ending in five because we hope in the next phase of the project to include census data, which is collected for years ending in zero. Thus we increase our temporal coverage.

2. Georeference the directory content to current and historical maps of the city and create a geographic interface for searching for and discovering directory information about and visual images of the city.

3. Demonstrate that it is possible to organize data by spatial location and then link this data to many other pieces of data that share the same location without having to extensively catalog them by subject, name, or other subjective terminology. And show that this method can lead to enhanced service for patrons, and more efficient use of resources for institutions. This third element will be more specifically explained in the Evaluation section of this application.

Building on the GIS and electronic text creation skills developed in the Edwin Bolles Archive on the History and Topography of London (an IMLS-funded project), we will use this project to extend the capabilities of the Tufts Digital Library GIS interface to include not only gazetteer information as in the London project, but directory information as well.

City directories continue to be invaluable resources for scholars, genealogists, and researchers who study for example: demographic trends, urban and environmental planning and renewal, ethnicity, immigration and migration. Yet, the information within

the directories is often difficult to extract and compare when working with the directories in their paper form. It is well understood that digitizing directories, dictionaries, encyclopedias and other reference type works allows researchers to more flexibility in how they can use and query these resources. Extending this flexibility to mapping the contents of the directories creates a visual picture of the data not previously available, and not practical to create, except through automated means.

This project supports and is related to the Boston Streets Photography Project, a collaboration between Tufts University and the Bostonian Society. The Boston Streets photo project (See Appendix D) will digitize and catalog (with geographic metadata), images of Boston from the collections of the Bostonian Society and Tufts University. That project is funded by a grant from the Fidelity Foundation.

The Bostonian Society has images from the earliest days of photography in the U.S. and covering all the city's many neighborhoods. . Tufts's image collection focuses on the neighborhood known as the South Cove, site of the current Medical and Dental School campuses.

The two projects will be tied together to allow for example, photo researchers, architectural historians, and urban historians to study the development of the city in maps and pictures, and to mine the wealth of demographic data included in city directories of the late 19th and early 20th centuries. The GIS-based system will be part of the structure for a union catalog, designed by Tufts as part of this project, that will create the possibility of a collaborative database of digital images of historic Boston street photos from a variety of institutions.

Tufts has significant experience in text and image scanning, markup, and cataloging workflow. The DCA now provides text and image scanning services to other libraries and departments at Tufts. The city directories present some interesting issues in text scanning and delivery.

First, the directories are large. The1865 directory has over 400 pages, with approximately 200 directory pages. As the years progressed, the directories significantly increased in size. The 1875 directory has 1200 pages, 500 of them devoted to directory information. The 1500 page 1885 directory has 1000 pages of listings, and by 1915 the directory has grown to 3000 pages with 2000 pages of directory listings. Managing text projects of this volume requires a good understanding of workflow and management.

A second area of interest is in copyright. A recent Supreme Court ruling clarified the fact that while directory information is in the public domain, value added material, such as advertisements, are subject to copyright laws. As a result, we will digitize the entire contents of the pre-1922 directories, but only the directory pages of the post-1922 volumes.

The basic workflow design for text creation is outlined below

Basic steps in text creation workflow process

Activity	Notes
1. Scan pages	Use sheet feed scanner where extra copies of the directories are available, or where disbinding is not possible, a planetary scanner purchased for the project.
2. Page image clean-up	Straighten and crop pages using Perfect Page Positioning software purchased for the project.
3. OCR	Prime OCR
4. Prepare uncorrected OCR and page-front image book	
5. Send scanned pages and uncorrected OCR of directory pages only for clean-up and minimal tagging	Clean-up and tagging based on our specifications, tagging for name (corporate or personal), occupation, street names, address.

GIS workflow:
The GIS center at Tufts has a base map of modern Boston. Historical maps will be scanned from the collection of the Bostonian Society. Although other institutions have scanned maps of Boston, none of them were scanned at the resolution or format necessary for our purposes. We are negotiating with the Library of Congress for appropriate formats of some of their scanned maps of Boston, and if obtained, will greatly increase the number of historic maps available. We will scan maps that most nearly match the years of the scanned directories. These maps will be oriented to the modern base map and overlaid on it.

In order for the system to work, we must create two things: a master list of geographic locations for our selected time periods, and a master list of addresses, business and personal names from city directories and photographs from our selected time periods.

We can use spatial data from the modern base map to georeference historical locations where appropriate. If a building, street or address exists on the modern map and the historic map, we can tie the geographic coordinates from the modern map to the historic map. With a georeferenced historic map, we can collect and assign geographic coordinates to buildings, streets, or addresses that do not exist on the modern map.

As a product of the manual georeferencing and the scanning and tagging of the city directories in the previous steps we can now assign geographic coordinates to all toponyms and addresses in modern and historic resources, (city directories, other documents, and images). From the tagged directories and image metadata we can automatically generate authority lists of addresses and business and personal names that

we can map to the master authority lists of geographic locations keyed to specific historical time periods.

With the London project we created the infrastructure to easily create new maps and data. The London project records changes over time at the city level. We can see the growth of the city over time in new streets, rearranging of old streets, landfills, open spaces, etc., but we don't know how this affects the people living and working in the city. In the Boston Streets Project, we are collecting and incorporating data that allows us to record demographic and commercial changes at the street level of a major city. For example, we can visualize the movement of the commercial district from the old downtown towards the Back Bay as the landfill project progresses; we can visualize the influx of Italian immigrants into the once Irish North End neighborhood; we can record the disappearance of 19th-century professions, such as shipwrights, from an increasingly industrialized city.

While the GIS interface is a tool to help visualize these historical events, the directories themselves are an alternate entry into the data, as are the photographs. All of these resources will be tied together through their association with a geographic space. It is this association that makes the whole more valuable than the individual parts.

The system is extensible, that is, more maps, images, directories or other media and documents can be added in the future. The city directories are important to start with, because they form the base reference tool to which other documents and media can be associated. Our plan is to extend the content by adding census information to the collection. Census data is even more richly detailed than directory information, yet still retains it strong geographic identity. If time and resources permit, we may import some census data during the time period of this project, and we certainly will as part of a phase II plan.

Management Plan
Tufts has a track record of successfully developing digital projects. The transformation in 2001 of the University Archives into the office of Digital Collections and Archives provides a stable home for digital projects that are integrated into the core mission of the university, and are supported over the long-term by the university budget. The DCA has a digitization facility that will include, as part of the Boston Streets project, a face-up planetary scanner for scanning bound volumes. Additionally, the Perseus Project, a digital library research and development organization is an ongoing collaborator with the DCA, and works with the DCA to implement practical applications of Perseus research. While the DCA is wholly in charge of the Boston Streets project, the expertise of the Perseus staff is always available to help troubleshoot technical problems should they arise.

More information about the management and project development experience of the staff can be found in the Personnel section of this application.

Budget

As is the case at many institutions, budget lines exist to support infrastructure, but funds for content creation are often the subject of grant proposals. Therefore the core of our request to the IMLS is for staff required to do the labor-intensive work of text preparation, scanning, and mark up. Tufts is supporting the equipment and infrastructure of the project.

Contributions
In the last year, Tufts has created the technological and administrative infrastructure to support and manage digital projects over the long term. The Office of Digital Collections and Archives was created out of the University Archives and permanent budget lines were created to support long-term storage of digital content using a central Network Storage Appliance, to obtain necessary equipment to support digitization projects, and to staff both the technical and strategic planning and development of the Tufts Digital Library. Tufts is also a partner in the Mellon-funded FEDORA repository project lead by the University of Virginia.

Additionally, part of the $10,000 Tufts share from the Bostonian Society collaboration will support map scanning and manipulation.

Personnel
Permanent staff:
Greg Colati, will be the project manager and budget officer. His is the principle investigator for the Edwin Bolles Archive on the History and Topography of London, an IMLS grant project for the years 2000-02. As Director of the office of Digital Collections and Archives at Tufts University, it is squarely within his mandate to develop digital projects. Mr. Colati will also coordinate the collaborative work between Tufts and the Bostonian Society.

Robert Chavez, who designed and implemented the London atlas, will oversee the development of the Boston atlas. Before becoming Technology Manager at the DCA, Mr. Chavez spent three years designing web-based geographic interfaces at the Perseus project (http://www.perseus.tufts.edu). Mr Chavez will design and implement the geographic interface, and work with the project assistant to georeference the scanned maps and ensure that specifications for outsourced work meet requirements.

The GIS specialist is a new position soon to be filled in the Tufts Academic Technology department. The GIS specialist will be the main link and point of contact between the project and the faculty who would incorporate it into their teaching. The GIS specialist will also work with the DCA staff to load modern base map of Boston into the Tufts DL system.

Temporary Staff:
Project Assistant, (to be hired). The project assistant will manage the day-to-day aspects of the scanning, tagging, and writing specifications for OCR outsourcing. The project assistant will supervise student assistants and perform quality control checking on their work. The project assistant will be primarily responsible for creation of the on line

directories. The project assistant will also be responsible for designing and implementing the project evaluation measures.

Student Assistants-20 hr/week--Student assistants, usually graduate students, or work-study undergraduate will do basic scanning tasks related to digitizing the city directories.

Project Evaluation
First, and most simply, we plan to implement a working geographic-based access tool that is used by both Tufts and Bostonian Society patrons to search for and request images from the collections.

Second, we hope to significantly alter the way photographic reference and reproduction is supported and supplied at the two institutions. We already know based on the 3000 images from the Tufts History collection already on line, that being able to deliver high-quality reference images has altered the resources devoted to photo reproduction orders, and the workflow of filling orders. We plan, with help from the Bostonian Society, to measure the effect of the new access tools on reference practices by compiling statistics and anecdotal evidence about how patrons discover and request photographs from the collections. We will also compare workflow procedures at the beginning and end of the project to discover the effect of the new system on cataloging and reproducing images.

 Third, we want to see GIS approach to research incorporated into the teaching methods of at least one class. To do this we will work students in the Tufts undergraduate research seminar taught by professor Reed Ueda, who will help us develop the public interface by using the system for research during the course of the seminar. We will work with the GIS specialist to use our experience with this seminar to introduce this tool to other faculty and students.

Finally, we would like at least two other institutions to consent to participate in a second phase of the project that would add collections from institutions beyond the two involved in this phase.

Dissemination
Both the Bostonian Society and the Tufts Digital Collection and Archives web site will feature the new interface. Additionally, we will seek to include stories about the project in the popular as well as scholarly press. For example, neighborhood group newsletters, local newspapers, etc. We will give presentations about the project at appropriate conferences where possible. We will visit other local institutions that have expressed an interest in the project and present demonstrations and solicit their participation in the next phase. We have already had significant interest in the second phase of the project from the Boston Public Library, the Society for the Preservation of New England Antiquities, and the Boston City Archives.

All modern cities are built up as collections of neighborhoods and districts whose boundaries ebb and flow with population changes and urban development and redevelopment. This presents a problem for anyone attempting to catalog and retrieve

images on the basis of neighborhood. In a related activity in the second year, we will begin discussions with other institutions in the area that hold similar collections on common metadata approaches to describing Boston photographs, and develop some working agreements among participants to develop some definitive approaches to describing the city's locations.

Sustainability
In 2001, Tufts created the office of Digital Collections and Archives (DCA) to centralize the development of the Tufts Digital library. This was a major commitment by the University to preserve its digital intellectual resources and to support and manage digital projects over the long term. The Office of Digital Collections and Archives was created out of the University Archives and permanent budget lines were created to support long-term storage of digital content using a central Network Storage Appliance, to obtain necessary equipment to support digitization projects, and to staff both the technical and strategic planning and development aspect of the Tufts Digital Library. Tufts is also a participant in the Mellon-funded FEDORA repository project lead by the University of Virginia.

A collateral value of the project is the partnership with the Bostonian Society. While larger universities have the infrastructure to build and sustain high capacity public access networks and long-term storage capability, the same cannot be said for many small and medium-sized historical institutions. The Boston Streets project is not only a project to demonstrate a new means of access to important content but is a model for inter-institutional cooperation between universities and local historical societies to preserve and create access to our cultural heritage