Philadelphia Area Consortium of Special Collections Libraries
c/o The Historical Society of Pennsylvania
1300 Locust Street, Philadelphia PA 19107
215-985-1445 voice, 215-985-1446 fax
http://www.pacscl.org/

Greater Philadelphia GeoHistory Network

Report to the Andrew W. Mellon Foundation
Planning Grant
January-December, 2005

Next Steps

We plan to develop the Greater Philadelphia GeoHistory Network on four fronts: 1) technology infrastructure; 2) historical data; 3) communication and networking; and 4) training, evangelism, and support. This section gives a detailed analysis of the technology and infrastructure issues and offers a summary of the planned approaches on the other three fronts to meet the needs of scholars and researchers. We conclude with a preliminary review of other potential organizational partners.

These next steps do not constitute a prospectus. We expect considerable refinement based on input from symposium participants and further discussion.

1) Technology Infrastructure Development

The first key component of the future GPGN needs to be a robust, scalable, and accessible technology infrastructure. The purpose of such an infrastructure is (1) to provide the tools necessary to locate places and addresses accurately in a given temporal context and to follow their change over time, (2) to support the creation of metadata describing both geography and time, documenting resources in institutional collections, and (3) to provide the necessary components that allow project developers and general end-users to locate resources (photographs, manuscripts, maps, census data, and directory information) across multiple institutions and partners, using geography and time as primary search considerations.

Conceptual Hurdles

Conceptually, there are significant challenges that we face in developing a technology infrastructure for the GPGN. In terms of metadata, we must agree on what constitutes the basic level. While significant work has been done on developing the "basic standards" of metadata, there is still some disagreement over the necessary levels of description, ranging from the

minimal and accessible approach of the Dublin Core Metadata Initiative[1] to the relatively complex and inclusive Visual Resources Association Core Categories.[2] Because of the wide variety of information and resources represented in PACSCL collections, we envision an infrastructure that supports an equally broad set of metadata standards, mapped to a basic Dublin Core schema.

Beyond that, we must agree on how to represent location and date, developing methods to appropriately represent the relative precision and certainty of such metadata. Dublin Core does not provide any standard methodology for such data; rather, it leaves that choice to the implementers of the standard. There are some standards in use, particularly by commercial software providers such as ESRI, that provide good location descriptors, but there are no existing standards that provide answers for the problems of precision and certainty, let alone in temporal context.

Much of the material in our collections can be approximately described in terms of geography – "5th and Chestnut Sts;" or perhaps, just "Chestnut St." or "near Philadelphia." Much of the provided metadata are also uncertain – information about photographs from scrapbook captions that are already known to contain errors. All of these provide conceptual challenges that have no cut-and-dried answer; development of one standard approach to documenting location, date, precision, and certainty is both necessary and achievable. Developing such a standard, however, will require new research, discussion, and agreement, and will probably require the involvement of various national cataloging communities. Developing a geographic location schema for metadata could be a major contribution to digital library development.

We must also determine how to approach the aggregation, correlation, and change of administrative units, place names, and addresses over time. Even within its relatively short history, Philadelphia has undergone significant change, incorporating adjacent municipalities, redrawing ward boundaries, and renumbering every address in the city. Developing methods for metadata creation and searching that respect not only the numerical geographic coordinates, but also the context of those coordinates in history, is essential.

Finally, there is the challenge of presenting all this information in a way that is usable and understandable not only by GIS professionals, but by scholars, researchers, archivists, librarians, and, eventually, members of the public. As part of the GPGN project, we expect to incorporate and expand existing applications (such as the City of Philadelphia's Photo Archives) by enabling them to participate and draw from the larger resources of the GPGN. We also expect to build some basic, innovative applications that provide collaborative geographic and date-based searching. These will serve as kindling for future projects that make more extensive use of emerging ideas of computer interfaces and human interaction in ways appropriate to various audiences.

We realize the focus of the GPGN cannot be "everything for everyone," but rather to provide the right tools and the right data to permit scholars to begin to parse, assemble, and analyze historical geographical information in ways that make sense and benefit research in Philadelphia and

---

[1] http://www.dublincore.org/documents/dcmi-terms/
[2] http://www.vraweb.org/vracore3.htm

beyond. The goal is to create a single platform for the melding and visualization of vast stores of historical data, resulting in revelations and clarifications and providing a catalyst for actionable insight.

Technology Model

The GPGN proposes gathering data from a long list of Philadelphia institutions (and other institutions that have data and resources about Philadelphia history), and using that data to build a few applications that present historical images, manuscripts, and data in geographic and temporal context. Looking past the specific data-sets and the specific institutions, there are two basic technical approaches to organizing a system such as this.

The first, which is easily understood and frequently implemented for projects such as this, is a **centralized model**, where all the data necessary for each application are gathered and stored on a central server (or group of servers), and are accessed through one or more user interfaces.
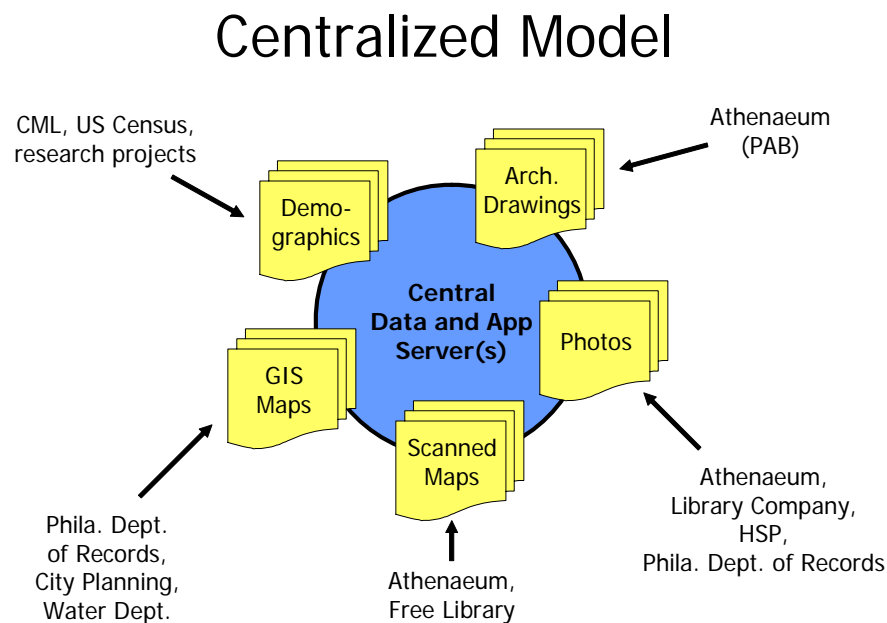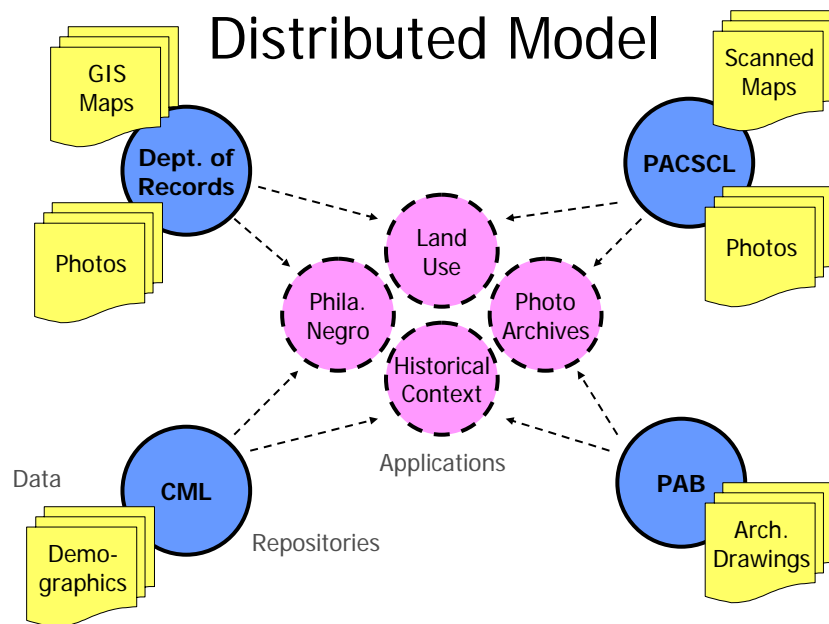
Centralized Model

In the centralized model, data for the applications are stored in one place. The data repository system is built to support the applications (or the clearinghouse, as the case may be), and is generally not suitable to be used also for the internal needs of the various contributing institutions. For example, if the Philadelphia Department of Records were to contribute their coverage of photographs in their holdings, the central system would be unable to accommodate internal data processes (ongoing entry and scanning of old and new photos) and specialized needs (printing and tracking of photo reproduction requests).

Contributing groups would have to submit to strictly-controlled metadata standards, and would have to figure out processes and policies for data refreshing, access control, and digital rights management. Already such things are difficult to agree on; gathering data and images together on a shared system may only increase those difficulties.

In addition, a centralized system is difficult to plan, scale, and host in a loosely-federated organization such as PACSCL. The amount of data may be overwhelmingly large, and each new institution would make additional resource demands upon the central system. This model may scale to support a large number of datasets, but it is unlikely that it would scale gracefully to support any extensive (multi-state or nationwide) collaborative projects.

The second approach is a generalized hybrid of the first, and this is the model we are investigating for the GPGN project. In a **distributed model**, data repositories are separated conceptually and perhaps physically from each other and from the applications that utilize the disparate data sources.

Data are stored and maintained in separate repositories (which may actually be centralized repositories for the respective organizations, fitting the first model above). This allows metadata to be maintained in one place, in repositories that support local institutional needs (such as providing photo reproduction services), and facilitates enforcement of local copyright and access restrictions. With distributed metadata and images, each institution or consortium is responsible for scaling the technology resources as necessary to support its own collections. No participant is a burden on any other participant.

The interactions between the repositories and applications are similar to the model developed by the Open Archives Initiative (OAI). Data in the various repositories can be maintained using metadata standards that are appropriate to each – with the various schemas mapped to a minimal set of descriptors (the Dublin Core elements). Each repository maintains a special gateway that responds to queries from applications or metadata harvesters and returns metadata from its own sources.

The model described here has already been demonstrated, and is seeing increased use as OAI support grows. Thus the primary challenge in building this network is incorporating into it the necessary services and standards effectively to capture geo-temporal information – to provide, exchange, and interpret accurate metadata for the geographic and temporal content and context of an item.

From a technical standpoint, the GPGN will function as a system of distinct yet complementary "services." Using data prepared for the project and the resources of data available from participating institutional repositories, these services will enable collaboration, cross-repository searching, and advanced applications, all built on a geo-temporal framework – that is, one that fully provides for spatial and temporal metadata. The sections below provide an overview of these various services and how they will relate to form the infrastructure of the GPGN.

Time-enabled Geographic Tools

The constant challenge of a standard Geographic Information System is that the physical world continues to change and the data are always one step behind. The concept of a Historical GIS adds, then, the complexity of gathering and using all of these geographic data over a wide span of time.

As do Library of Congress subject headings, so also place names, and the terms used to describe place, change significantly over the years. As a result, it is difficult, if not impossible, to search for information successfully based on geography in current systems. A search using the current name may fail to return information from a potentially significant source referring to the place by a name used fifty or two hundred years ago. A **gazetteer** service will provide name, coordinate, and boundary authority for places in historical context. Initially, we will rely on existing work to create the basic set of data to support such a service; moving forward, the system will be designed to permit individual researchers to submit updates to be reviewed and incorporated into the database. This then builds a collective knowledge about place that traditionally has resided only in a few local experts.

Even with an accurate historical gazetteer, the ability to locate information finely in dense urban areas is very limited. A **historic street and address database** would link addresses from the past into their respective geo-referenced counterparts in the present. GIS data coverages of historic street centerlines can begin to be created from existing electronic data sources. This can be refined by digitizing and incorporating street data (opening dates, address number ranges, geographic coordinates) from the Philadelphia Streets Department, and by comparing real estate atlases at key points in the city's history to determine number mapping before and after consolidation of the city with the county and the address numbering changes in the mid-1800s. As with the gazetteer, collaborative features will permit continued updating of the historic street and address data.

This is the basic toolset – the *sine qua non* – of a collaborative GeoHistory effort bringing together diverse repositories. These tools, collectively, provide a "geo-temporal naming authority" in a field currently devoid of a standard. While the data collected for the GPGN will document the Philadelphia region and the city in particular, the standards, methods, and software

to be developed will be appropriate for any area. Also note that we fully expect the data collection aspects of building these tools to be more time-consuming and difficult than the development of the software components. This highlights the importance of the ability for scholars to contribute as their research uncovers additional information.

Geo-temporal Metadata Tools

Given a gazetteer and a database of historic streets and addresses, the next requirement is to develop the tools necessary to apply the naming authority to specific metadata records. The first of these is a **geocoding tool**, which will perform standardization on place names and addresses and assign geographic coordinates. While geocoding tools already exist and are widely used, this **temporal geocoder** will incorporate the historical context in providing a standard reference and in determining geographic coordinates. Besides the ability to geocode a single reference, organizations will be able to submit lists of addresses to a **batch temporal geocoder** and receive a database with geographical coordinates assigned. These tools will be used during the creation or augmentation of metadata describing items in various collections. They will be most useful for items with precise location and date.

In addition to items that have geographic context, scanned historical maps are some of the basic building blocks of the project and are intrinsically useful. Their value is enhanced even further by the ability to align the maps with maps of the same geography from other times. This allows overlaying of multiple maps for a clear visualization of the morphing of characteristics over time. This can be accomplished with existing technology, but has been beyond the capabilities of most institutions due to the expense and complexity of GIS software. With the development (as part of the GPGN) of an **image registration tool** focused specifically on rotating, scaling and moving a map image to a location on a base map, it will be feasible for contributors to the GPGN to create not only metadata describing the coverage of the map, but also to link points on the map to geographic coordinates, allowing the map to be projected in a fashion such that each point gains geographic significance. When users are searching for a specific location or set of coordinates, the system will then not only be able to identify which maps contain the requested points but where on that map the points lie.

Distributed Geo-temporal Search Matrix

As with the metadata creation tools described above, the basics of the "search matrix" focus around data and maps. The services in this case will be provided on a centralized website, but also as part of the various repositories contributing to the GPGN.

In terms of metadata, the search matrix will consist of **metadata providers and harvesters**, and will be built similarly to the model developed by the Open Archives Initiative. The additional functionality required in the GPGN is the ability to search and filter results by geography and time. The standards developed for representing geo-temporal information will be incorporated into these systems, permitting distributed and federated access to the various repositories contributing to the GPGN. While much of this work happens behind the scenes, the GPGN will also provide a **human user interface** to search this matrix and link to the various resources discovered.

While the first service is to provide distributed access to metadata and the collections of contributing institutions, the second service is to provide **contextual historic maps**. Generation of base maps will be possible using not only the standard geographical parameters of extent, size, scale, and layer visibility, but also the qualifier of time. Thus, a user will be able to generate a map of 1880 streets, for example. Improvement in historical street name and centerline data will enable the display of maps based on a particular year.

Project and Data Repository

One of the constant dangers of electronic data and technology projects is that they frequently get lost – forgotten or no longer accessible or usable due to changes in technology, failures in maintenance and storage, and poor documentation. In addition to enabling users of the GPGN to explore and use the resources of traditional collecting institutions, we must also provide resources to facilitate discovery of prior work. Much of the value of geographic and historical information comes from research and synthesis; ignoring this work of the scholarly community in building an infrastructure such as the GPGN would be unthinkable.

Thus the GPGN will include a **project repository** that will accept submissions from researchers as to the status, scope, and location of their work. The project repository will not necessarily seek to physically store the data of the various projects documented, instead relying on existing organizations such as the Pennsylvania Spatial Data Access (PASDA) group at Penn State. PASDA has been successful at serving as a repository for the "primary source" data produced and contributed by agencies across the state. Working with them we will seek to expand that mission to capture "secondary source" project data and develop methods of documenting and searching this material.

Data Maintenance and Migration

The digital library community has devoted considerable thought to the maintenance and migration of data as technologies evolve. Definitions of and standards for "trusted repositories" have been developed by the digital library community, and as GPGN develops its plan we will have to agree on how our members will implement and abide by these standards. Without such agreement, the integrity of the GPGN will be difficult to ensure.

2) Data Development

At the symposium, participants were encouraged to think of the kinds of historical projects the GPGN might embrace, offering us some guidance from the "demand" side. In addition to being unanimous in their desire for historical geo-rectified maps, they offered some potential research questions to be addressed with the assistance of the GPGN, as identified during the course of the symposium. This list is not exhaustive but illustrates a range of thinking.

- Evidence of city planning, formal or informal, from the eighteenth to the mid-twentieth centuries

- A look at natural resources in the region and the ways in which they influenced residential, agricultural, and industrial development

- A racial and ethnic geography of the city over time in relation to industry, the development of labor, and health (a historic epidemiology of the city)

- A historical geography of Philadelphia's economic and cultural rise in relation to other cities

- An animation of land use change and urban growth

- A historical GIS of the Civil War, for example, looking at such issues as

  o "in and out of Philadelphia": the flows of supplies and people
  o The impact of labor loss and shifts of labor onto women and children
  o The impact of mortality on the home front – were certain neighborhoods suffering more loss of menfolk than others?
  o Donation of funds to the war
  o The growth of war-generated wealth in industries such as iron

- A historical atlas of Philadelphia

- Intellectual lineages (plotting the impact of a particular scholar or researcher as reflected in the geographic distribution of his/her students)

From the "supply" side, some collections are relatively accessible and relevant to these research questions and can be early targets for inclusion, namely:

- Additional historic maps, such as real estate atlases, or the hundreds of maps of portions of the city prior to its 1854 consolidation – including additional work on the additional plates scanned in the planning project and contributed from a complementary project for the Free Library of Philadelphia.

- Graphics collections, such as the 2,500-plus Hexamer & Locher real estate surveys, the 1,500 D. J. Kennedy watercolors of sites in Philadelphia, 1860-1890, and selected items from the 2,000,000 photographs in the City of Philadelphia archives.

- Data sets, such as the historic street name index, city directories, building permits, and administrative units (maps of various townships, boroughs and wards of the city since its founding), and census data.

In addition to these relatively accessible maps, images, and data sets, we are looking at ways that other resources in PACSCL member collections may be located and linked to the GPGN. In an initial survey of members that is serving as a pilot for data collection, we asked for information on collections relating to the 700 block of Chestnut Street; this survey met with limited success

because, as we now recognize, our members are not accustomed to indexing materials geographically. A follow-up survey, which is in preparation, will identify prominent individuals or institutions tied to this block of Chestnut Street and will query members for collections relating to these entities. Our experience with this second survey will suggest whether or not this approach is scalable. The results of our second survey will also engage us in solving the problems of developing metadata to represent location and time.

Other sources of data to be uncovered in a systematic fashion and incorporated into the GPGN include the collection level records to be developed as part of the Mellon-funded Consortial Survey Initiative, and, potentially, from the statewide digitization initiative (both described in our final report, page 7). Finally, we are looking to build in communications mechanisms (see below) that will help our selection of resources to continue to be demand-driven as a complement to a phased and prioritized approach to data selection and inclusion. As noted in the section of the report on the symposium (part 3 of the final report), several attendees were eager to find ways to involve data sets ranging from census data to a survey of nineteenth-century industrial sites.

3) Communication and Networking

The planning project, particularly the symposium, demonstrated that there is a large and growing community with interest in historical GIS, including both those who have collections and the scholars who will use them. Our Advisory Committee reflected this diversity and should be expanded in any implementation phase, perhaps creating a GIS Roundtable of the key dispensers and users of geographical information. Another invaluable tool in the development of communication on this subject is the website itself, which has already proven to be useful as a clearinghouse, a means of announcements, and a collector of survey information.

Our plans as we go forward, which will help with the creation of our prospectus and eventual proposal, include encouraging more interactivity among symposium participants and other members of their professional network. As a first step, we are e-mailing attendees to tell them about the availability of symposium "product" – Power Point slides, summaries of talks and of breakout discussions. At the same time, we are inviting them to opt-in to a mailing list that will allow us to share information on public programs, new resources, and new techniques. We will seek to engage them as networkers on behalf of the project, both forwarding information and encouraging others to join the e-mail list. We are considering other means of communications enhancement, such as a blog, a bulletin board, or a Wiki, to allow information sharing in a collaborative environment.

Lastly, it was universally acknowledged that future symposia of the sort hosted in 2005 would be enormously beneficial to the community at large.

4) Training, Evangelism, and Support

Technology, data, and personal networking are excellent at making significant resources available to the community of professional geographers and GIS users. But these fall short of

achieving one of the primary goals of the project, which is to make these resources available to scholars who are currently outside of the disciplines that use GIS extensively.

The three pieces necessary are 1) *training,* providing necessary skills and resources to those who are interested in using GIS technology in research; 2) *evangelism,* seeking to expand understanding of the GPGN and its resources in the scholarly community; and 3) *support,* providing conceptual, technical, and collaborative assistance for various projects that propose to make use of the GPGN.

What we envision is a local model akin to that employed by NITLE, with a focus on using GIS technology in research as opposed to integrating it into curricula. We have partners who are capable of helping to build this network. The Cartographic Modeling Lab (CML) at the University of Pennsylvania is well-suited to training, evangelism, and support. Its focus to this point has been on the use of current GIS data for land use analysis and planning, community development, and historic preservation. Implementation of the GPGN could see the CML expand its role into providing resources for historical GIS work. In particular, CML could combine its experience with training non-technical contributors, e.g., staff at community development corporations, with NITLE's experience in the liberal arts field. The software vendor ESRI's extensive experience with user groups, both geographic and topical, is an additional resource, and ESRI has already indicated its eagerness to partner with this initiative by its support of and participation in the symposium.

Partnership with NITLE in parts of this process will be a significant boost to our efforts, as they are already engaged in work with various Philadelphia-area institutions. We expect to pursue this relationship as plans for the GPGN progress, beginning with some training sessions that Diana Sinton, head of NITLE's GIS initiative, is scheduled to conduct at Bryn Mawr College later this year. We would also like to engage in discussion with the Council on Library Information Resources (CLIR) in conjunction with a project team presentation in April.

Lead Organizations

Three key collaborators were originally identified for participation in the GeoHistory Project; PACSCL, the City of Philadelphia Department of Records, and the Athenæum of Philadelphia. Since the beginning of the pilot project, each of these experienced growth in institutional capacity that positions them to remain central.

- PACSCL is conducting a Mellon-sponsored survey of under-processed collections within its member institutions. Coincidentally, these collections are often geographically rich. A coordination of the implementation of the GeoHistory network, with new processing and cataloging protocols, will strengthen both efforts and will add an additional tool to the unified catalog searching, using the Z39.50 interface, currently under development as part of another initiative of the Greater Philadelphia Research Collections Network.

- The City of Philadelphia is nearing completion of a $6 million project to create seamless coverage for its historical base map and has also launched its photo archive website,

which has thousands of photographic images digitized and available for public viewing. The City is committed to coordinating its efforts with the creation of the GPGN.

- The Athenaeum has added more than 60,000 records to the Philadelphia Architects and Buildings Database since this planning grant began, including 15,000 fire insurance surveys for Philadelphia buildings, which provide address, tenant, and building use information. In addition, the Athenaeum's Regional Digital Imaging Center (RDIC) will continue to play an important role in the safe scanning of oversize maps and related materials.

In addition to these three original partners, two others need to be considered for inclusion in the planning and implementation of the GeoHistory Network at this time. The first of these is the Cartographic Modeling Lab (CML) of the University of Pennsylvania, which has played a key role in using current technology to deliver geographic information to a scholarly audience and beyond. Its strong ties with the academic community are particularly useful, and a new degree program, Masters of Science in Urban Spatial Analysis, began during our Mellon-funded planning grant. CL has been an unstinting supporter of our project from the start. Its key potential role in training and evangelism has already been noted.

Next, Avencia, Inc. under the leadership of Robert Cheetham has played a major role in the conceptualization of this project, both in the planning stages and, especially, in the vision for its next steps. Its efforts have included the building of software to manage and search both contemporary and historic cultural resources in Philadelphia, including the City's Photo Archives, a Unified Land Record System, Historic Street Names, and a hand-held "Pocket Culture Browser" for tourists as well as many applications at CML, including their Neighborhood Information Base, MuralBase, and CrimeFinder functions. Avencia's participation in any future GPGN efforts will be invaluable.

As we progress with our plans in the areas of communications, networking, training and evangelism, it is entirely possible that some of the organizations we now see as contributing members will evolve into lead organizations. The initiative has the potential to combine a rich set of organizational players to support a rich set of historical geospatial resources working together to make Philadelphia's history and environment geospatially accessible to a greater degree than that of any other American city.