NATIONAL ENDOWMENT FOR THE
Humanities

OFFICE OF **DIGITAL HUMANITIES**

Narrative Section of a Successful Application

The attached document contains the grant narrative and selected portions of a previously funded grant application. It is not intended to serve as a model, but to give you a sense of how a successful application may be crafted. Every successful application is different, and each applicant is urged to prepare a proposal that reflects its unique project and aspirations. Prospective applicants should consult the Office of Digital Humanities program application guidelines at http://www.neh.gov/grants/odh/digital-humanities-start-grants for instructions. Applicants are also strongly encouraged to consult with the NEH Office of Digital Humanities staff well before a grant deadline.

Note: The attachment only contains the grant narrative and selected portions, not the entire funded application. In addition, certain portions may have been redacted to protect the privacy interests of an individual and/or to protect confidential commercial and financial information and/or to protect copyrighted materials.

Project Title:	World-Historical Gazetteer
Institution:	University of Pittsburgh
Project Directors:	Patrick Manning (University of Pittsburgh) and Ruth Mostern (University of California, Merced)
Grant Program:	Digital Humanities Start-Up Grants, Level 1

Table of Contents

List of Participants

Project Participants	**Institutional Affiliations**
Patrick Manning*	University of Pittsburgh
Ruth Mostern*	University of California, Merced
Merrick Lex Berman	Harvard University
Peter K. Bol	Harvard University
Kai Cao	University of Pittsburgh
Matt Drwenski	University of Pittsburgh
Tom Elliott	New York University
John Gerring	Boston University
Karl Grossner	Stanford University
Leif Isaksen	Southampton University
Evgeny Karataev	University of Pittsburgh
Matthew Knutzen	New York Public Library
Rainer Simon	Austrian Institute of Technology
Humphrey Southall	Portsmouth University
Vladimir Zadorozhny	University of Pittsburgh

**Co-Project Director*

GRANT11484019 -- Attachments-ATT2-1235-participantslist.pdf

Abstract and statements of innovation and humanities significance

Abstract

This project will advance work toward creation of a world-historical gazetteer that will provide comprehensive databases of places throughout the world since 1500 CE, including attention to the range of attributes known for each place. To satisfy the needs of all the large-scale historical data resources now being created, there is need for such a comprehensive and general gazetteer system. The convening of a two-day workshop, including leading figures who have developed gazetteers and the datasets in which they are incorporated, will bring about a research design for this world-historical gazetteer system, which can then be implemented in subsequent work. Four small research tasks concerning services, standards, and content will bring immediate advance toward implementation. The project is organized by the Collaborative for Historical Information and Analysis (CHIA), which has a record in sustaining collaborations for large-scale humanities work.

Statement of Innovation

In organizational innovation, this project gathers many of the leading contributors to spatial analysis over time, to minimize duplication of effort by developing common standards for a system of gazetteer services. In analytical innovation, it addresses the full range of types of spatial documentation, from the highly precise to the significantly uncertain, at scales from local to transregional—instead of addressing separately the various parts of this big problem.

Statement of Humanities Significance

The project will expand the frame of studies in history and other discipline with consistent databases of places over time and space. It will facilitate systematic identification of places worldwide, permitting new sorts of connections for fields in the humanities and related social sciences. The gazetteer will facilitate the organization and analysis of tabular data and provide a platform for documenting texts and images by place and time, advancing cultural as well as social scholarship.

GRANT11484019 -- Attachments-ATT3-1236-abstract.pdf

Narrative

Digital historical gazetteers – databases of place names that include information about location, other attributes, and changes over time – are the foundation of spatial humanities reference and infrastructure. This Level I startup proposal seeks funding for a project to develop **content**, **standards** and **services** for a world historical gazetteer to serve as an authoritative resource for named places and administrative systems from 1500 CE to the present day. As a starting point, the gazetteer will include all historical place entities from the WikiData project and from GeoNames, and will be augmented by any new historical places submitted by CHIA partners. Our aim is to create a spatial entity reference system that is global in scope and is implemented as an open, freely accessible web service to support ongoing research by historians and social scientists.

Enhancing the humanities through innovation. Creating a world-historical gazetteer builds on the previous and current comprehensive work in gazetteers, notably that of the Pelagios 3 project (led by Leif Isaksen—this project focuses on the world before 1500) and the PastPlace project (led by Humphrey Southall). The world-historical gazetteer, which may well take the form of a distributed model rather than a single gazetteer, takes the further step of addressing a fuller range of problems. Places are described in a typology by a range of attributes. At one pole are places defined by location and boundaries with clear times of existence; at another pole are places with names but no precise location or limits on their time frame. In both cases, administrative units are distinct from places defined by communities, or places as un-typed entities. For world-historical datasets, which are to include data for the whole world despite the varying degrees of certainty, there is a need for gazetteers that flexibly include a wide range of spatial attributes; in addition, a gazetteer system will require description of exceptional cases.

The system will be integrated with the world-historical data archive that is under development by the Center for Historical Information and Analysis (CHIA), so it will permit spatially aware analysis of historical quantitative social statistics. It will also be designed as a free-standing work of reference, infrastructure and standards specifications for any project or user in the spatial humanities. The contents, in addition to being accessible from an open application programming interface (API), will be exposed in standard resource description framework (RDF) format to support linked open data. At present, no single gazetteer database or system of gazetteer services allows for historical named-entity query, visualization, and management at the world historical scale. While much development in this domain has occurred over the last twenty years, there are no consensus-based, community standards for design or services, let alone accessible global content. The purpose of this grant is to rectify those problems and to begin developing the needed content.

We seek start-up funding to inaugurate this project by completing two essential, integrated first-step activities. The first is a two-day working meeting in August 2014. This will be held at the University of Pittsburgh in conjunction with the annual CHIA meeting. We are proposing a meeting because there are a number of promising historical and thematic gazetteer projects and linked open-data gazetteer projects presently underway—inside and outside of academia, and inside and outside the humanities—but the developers and scholars involved with them are only in loose communication. Moreover, many tasks necessary to global historical gazetteer development are not within the purview of any of them. The working meeting will include representatives from all of the major projects. It will ensure that subsequent CHIA world-historical gazetteer development is aligned with best standards and best practices, that it builds on related efforts rather than duplicating them, and that it enriches and coheres the vibrant intellectual community around historical and humanistic place-name entity scholarship and development. The meeting will inventory and analyze existing regional historical gazetteer content development projects and gazetteer infrastructure and standards projects; it will reach firm decisions about subsequent steps. For the workshop agenda, see Appendix A.

In addition to the meeting, we are funding a Pitt history graduate student, Matt Drwenski, to support four small development tasks. He will assist in building gazetteer capacity into the CHIA information system (**services**). He will also assist in providing a working implementation of a historical place name API that includes temporal queries and a draft proposal for a temporal gazetteer schema for

GRANT11484019 -- Attachments-ATT4-1237-narrative.pdf

interoperability between gazetteer services (**standards**). Finally, he will assist in building a testbed of global scope. This involves two tasks. One is to ingest and standardize the 1.4 million entries in English-language Wikipedia with geographical coordinates according to the requirements of this proposal. The other is to identify gaps in coverage, supplementing the Wikipedia testbed as needed by names harvested from print historical atlases. These development activities will allow us to complete the highest priority development tasks pertaining to the global historical gazetteer, and they will ensure that the working meeting focuses on results rather than abstract discussion.

This grant would solve urgent development bottlenecks for historical quantitative social science repository development, and it would permit several activities of fundamental significance for the spatial humanities: 1) minimize duplication of effort while raising awareness of best practices throughout the historical spatial-entity research community, 2) identify specific global times and places for which historical spatial entity data is poorly or unevenly documented and develop strategies to fill those gaps, 3) solve the technical and semantic problems associated with multiple attributions for historical spatial entities, which result in many kinds of overlapping, inconsistent and fluid kinds of data being associated with the "same" place.

History and duration of the project. For the participating groups of gazetteer developers, consistent work goes back more than a decade. Pelagios, a consortium of such groups focusing on the Ancient World, has worked in concert for over five years. CHIA took form in 2011, formally linking multiple institutions in the development of a world-historical data resource; the collaboration underlying CHIA had solidified by 2009 through joint applications for funding of a global-historical gazetteer.

Environmental Scan. This proposal builds upon and is integrated with numerous existing initiatives and research directions. Appendix B includes links, personnel and descriptive information about the projects listed here:

1. Developments in regional and thematic historical gazetteers. There are, at this point, a great many distinct historical gazetteers of various size, sophistication, and scope, covering parts of Asia, North America, and Europe. However, there are many gaps in coverage, and the structure of some projects is idiosyncratic.
2. Linked open data historical gazetteer development. The Mellon-funded Pelagios 3 project (including some of the same key personnel as this proposal) is the domain leader, but their work is nascent and is limited to pre-1492 content.
3. An edited volume about theory and practice for historical and "rich-data" gazetteers, *Placing Names*, edited by Merrick Lex Berman, Ruth Mostern, and Humphrey Southall, is under contract with Indiana University Press. The book is based upon gazetteer workshops at the Association of American Geographers in 2010 and 2011, which laid useful groundwork and solidified collaborations, but did not result in tangible development plans.
4. There are two existing gazetteer systems, GeoNames and Wikidata, that could reasonably serve as potential backbones for a historical gazetteer. While they are contemporary in scope, the GeoNames edit tool for alternate names allows any name to be flagged as isHistoric, and additional historical information such as dates or source citations could be stored separately and linked to GeoNames. Wikidata is a new initiative to create a structured and open data version of Wikipedia content. While we need to monitor it closely since it has launched so recently, we consider it the most likely base gazetteer candidate because of its size, extensibility, and the number of historical names it holds.
5. The Library of Congress (Schuyler Erle and Shekhar Krishnan of Topomancy.com) has developed a tool for conflation of GeoNames, which is critical for disambiguating and relating spatial entity information (http://gazetteer.in).
6. The Center for Historical Information and Analysis has received funding from the National Science Foundation to develop historical quantitative social data at the global scale: a world-historical data resource, and to develop a data ingest system (presently known as Col*Fusion). CHIA also has ongoing institutional support from the University of Pittsburgh and a partnership with the WorldMap and Dataverse projects at Harvard for spatial data visualization and data

GRANT11484019 -- Attachments-ATT4-1237-narrative.pdf

archiving. CHIA is the parent project of this proposal, and is committed to development, project management, fundraising and data management beyond the term of the start-up grant. There is additional detail about CHIA and Col*Fusion in **Appendix B.**

Work Plan. We plan to hold a working meeting at the University of Pittsburgh in late summer 2014. The agenda of the meeting is fully specified in Appendix A. During the subsequent year, graduate student Matt Drwenski will work under the direction of Patrick Manning and in consultation with workshop participants (notably but not limited to Berman, Southall and Zadorozhny) to meet the **content**, **standards** and **services** development goals specified above, such that by the end of the grant term, we will have completed the following tasks:

1. Linking the Wikidata gazetteer and the Library of Congress name disambiguator with the CHIA Col*Fusion data ingest system as a spatial named-entity reference guide for contributors who submit data to CHIA; also developing workflow in Col*Fusion for adding new place names from CHIA data contributors to the CHIA gazetteer. The **development** and **content** activities align with this task.
2. Inventorying existing gazetteers of national, regional and global scope, and coverage of the last five hundred years. We will complete a report of their temporal, spatial and thematic coverage, API standards and linked data capacity, database format, project developer, and state of completion, and assessing the feasibility and challenges of linking each one to the CHIA system. The historical API implementation work aligns with this **standards**-related task.
3. Finalizing a strategy to develop limited new content for a high level global-historical gazetteer where needed spatial entities are not covered in existing digital historical gazetteers. We envision new development of a few thousand entities including watersheds, regimes, second-order divisions (provinces), and national and provincial capitals for the last five hundred years, with priority for entities that are likely to be associated with quantitative data, and with print historical atlases the most likely source of content.
4. Identifying appropriate potential funding sources for further development of the gazetteer. Such agencies as the National Science Foundation, National Endowment for the Humanities, Andrew Mellon Foundation, and the Joint Information Systems Committee (UK) have already invested in development of digital historical resources, and may have interest in supporting this comprehensive advance in geospatial descriptive data.

Staff. The two Co-Project Directors are dataset and gazetteer developers and researchers with a long track record of success and collaboration.

The workshop invitees will additionally include CHIA research associates Kai Cao and Evgeny Karataev, CHIA board member John Gerring, Pelagios investigator Leif Isaksen, historical geo-semantics expert Karl Grossner, PastPlace developer Humphrey Southall, China Historical GIS architect Merrick Lex Berman, Center for Geographic Analysis director Peter Bol, New York Public Library map specialist Matt Knutzen, Tom Elliott of the Institute for Study of the Ancient World, Linked Data specialist Rainer Simon, and Col*Fusion developer Vladimir Zadorozhny. Matt Drwenski, in addition to participating in the workshop, will continue as researcher and liaison in post-workshop development work

While they are not part of the project staff, we hope for the opportunity to consult with such leading developers as WikiData developer Katie Filbert, GeoNames developer Marc Wick and Library of Congress conflation tool developer Shekhar Krishnan.

Final product and dissemination. The final product will consist of (1) a set of standards and criteria for worldwide documentation of places, (2) a database including a large quantity of consistently formatted specific data on places, drawn from research of participants and from other resources, (3) a system of ingest for continuing incorporation of additional gazetteer data, and (4) a user interface enabling researchers to apply the gazetteer. Dissemination of results and notification of tools and services will take place through the CHIA website (www.chia.pitt.edu), the *Journal of World Historical Information* (jwhi.pitt.edu), the *Placing Names* volume, and through direct communication with the gazetteer development community.

GRANT11484019 -- Attachments-ATT4-1237-narrative.pdf

Biographies

Merrick Lex Berman manages Chinese and Japanese geographic information projects such as China Historical GIS, the G. W. Skinner Regional Systems Data Archive, and collections of GIS datasets for modern China and Japan. In addition to developing placename databases, web services and digital collection tools, Lex works on GIS Data Portals, geographic ontologies, and a curriculum for Quantum GIS training.

Peter Bol is Charles H. Carswell Professor of East Asian Languages and Civilizations and the founding director of the Harvard Center for Geographic Analysis. He leads the China Historical GIS project, which is creating an historical GIS for China's history from 221 BCE to 1911. His recent publications include: "On the Cyberinfrastructure for GIS-Enabled Historiography," Annals of the Association of American Geographers (2013): Published online: 06 Jun 2013; "What is a Geographic Perspective on China's History?" in Chinese History in Geographical Perspective, eds. Yongtao Du and Jeff Kyong-McClain (Lantham: Lexington Books, 2013), pp. 197-204; "On an infrastructure for historical spatial analysis," Perspectives on History 50.7 (2012). "GIS, Prosopography, and History." Annals of GIS 18.1(2012): 3-15.

Kai Cao is postdoctoral associate at the World History Center, University of Pittsburgh, and Visiting Research Fellow at the Center for Geographic Analysis, Harvard University. He obtained his Ph.D. degree in Geography from the Chinese University of Hong Kong. Before joining the World History Center, he worked in research positions at the Center for Geographic Analysis at Harvard, the Department of Geography, University of Illinois at Urbana-Champaign, and Department of Geography, University of California – Santa Barbara. His research interests center on the applications of GIScience in social sciences and humanities.

Matt Drwenski is a graduate student in the Department of History at the University of Pittsburgh, specializing in world history. He works with his advisor, Patrick Manning, on research for the Collaborative for Historical Information and Analysis (CHIA). He received a BA degree in Political Science and History at Rice University (2008), in a program that developed his skills in interdisciplinary and quantitative analysis. He devoted the intervening three years to teaching history (especially AP World History) in a Houston public charter school.

Tom Elliott is Associate Director for Digital Programs at the Institute for Study of the Ancient World, New York University. He is leader of the Pleiades project, which is developing an online workspace for ancient geography. He previously served as founding director of the Ancient World Mapping Center at the University of North Carolina – Chapel Hill.

John Gerring is professor of Political Science at Boston University, specializing in comparative politics, methodology, and U.S. political history. His 2007-2010 NSF grant on the global impact of colonialism led to the Colonial Legacies project, now collaborating with CHIA. He is author of *Case Study Research: Principles and Practices* (2007) and *A Centripetal Theory of Democratic Governance* (2008) with Strom Thacker.

Karl Grossner is a geographer (PhD 2010, U.C. Santa Barbara) presently working at Stanford University Libraries as a Digital Humanities Research Developer. In that position he collaborates with faculty researchers, providing research design and software development support in intensive year-long engagements on selected projects (recently, orbis.stanford.edu; citynature.stanford.edu). The work is centrally concerned with spatial and spatial-temporal knowledge representation and analysis, and bringing digital tools and methods to humanities scholarship. Karl is a co-Chair of the GeoHumanities Special Interest Group of the Association of Digital Humanities Organizations (ADHO).

GRANT11484019 -- Attachments-ATT6-1239-biographies.pdf

Leif Isaksen is a Lecturer in Archaeology and Deputy Director of the Web Science Doctoral Training Centre, University of Southampton, UK. His research exploits the spatial aspects of historic materials in a variety of ways, through projects such as Pelagios, Google Ancient Places and Hestia. He is particularly interested in how digital expressions of place can be used to complement the spatial turn in the humanities. He is also undertaking research on the original global gazetteer - the Geographike Hyphegesis of Claudius Ptolemy. His publications include Isaksen, L., Barker, E. T. E., Kansa, E. & Byrne, K. Google Ancient Places: A NeoGeo Approach to Classical Resources. Leonardo, 2012, 45(1); Simon, R,, Barker, E. & Isaksen, L., 2012. Exploring Pelagios: A Visual Browser for Geo-tagged Datasets. In: International Workshop on Supporting User's Exploration of Digital Libraries, Paphos, CY, 23 -27 Sep 2012. pp. 29-34, and Isaksen, L. 2011. Lines, Damned Lines and Statistics: Unearthing Structure in Ptolemy's Geographia. E-Perimetron 6(4). pp. 254-260.

Evgeny Karataev is a doctoral student in the School of Information Sciences at the University of Pittsburgh working with Professor Vladimir Zadorozhny. His research interests include data management, crowdsourcing, and the combination thereof. He is currently working on the collaborative data integration architecture Col*Fusion.

Matt Knutzen is Geospatial Librarian of the New York Public Library's Lionel Pincus and Princess Firyal Map Division where he oversees, curates and ensures the long-term preservation of 453,000 maps, books and atlases. He helps NYPL envision new roles for networked map libraries, through map digitization and public engagement through tools designed to transform maps into more useful data.

Patrick Manning (manning.pitt.edu) is Andrew W. Mellon Professor of World History and director of the World History Center at the University of Pittsburgh. Trained as a specialist in the economic history of Africa, he has become a specialist in world history overall. His research has focused on demographic history (African slave trade), social and cultural history of francophone Africa, global migration, African diaspora as a dimension of global history, and an overview of the field of world history. Manning now directs the Collaborative for Historical Information and Analysis? (CHIA), in which five institutions collaborate under a grant from the National Science Foundation. His new booklet appears in 2013: *Big Data in History: A World-Historical Archive.*

Ruth Mostern (faculty.ucmerced.edu/rmostern) is an Associate Professor of History at the University of California, Merced, where she is the chair of the Interdisciplinary Humanities Graduate Group and the co-Chair of the Spatial Analysis and Research Center. She is a specialist in imperial Chinese spatial history and spatial history methods. Her book *Dividing the Realm in Order to Govern: The Spatial Organization of the Song State (960-1276 CE)* was published by Harvard University Press in 2011. Her co-edited book about digital gazetteers for history and the humanities, entitled *Placing Names: Enriching and Integrating Gazetteers*, is in progress and under contract at Indiana University Press. Mostern is a principal investigator on the National Science Foundation grant to the Collaborative for Historical Information and Analysis.

Rainer Simon is a research scientist in Computer and Information Science at the Austrian Institute of Technology, with special interest in Linked Data and digital preservation. Among his recent publications is "GeoPointing: Evaluating the Performance of Orientation-Aware Location-Based Interaction Under Real-World Conditions," co-authored T. Frohlich and T. Grechenig, in *Journal of Location Based Services* (2008)

Humphrey Southall's work centers in the construction of the Great Britain Historical GIS, a major resource bringing together historical statistics, texts and mapping. Total funding has been well over £2 million, and construction of the system has required the development of new methodologies. The

GRANT11484019 -- Attachments-ATT6-1239-biographies.pdf

resulting system has been used to support research in many fields, including health care, environmental management and Southall's own work on long-run socio-economic change in Britain, and the development of the north-south divide. The GBH GIS is also a popular resource for local history researchers, accessed via the web site A Vision of Britain through Time.

Vladimir Zadorozhny is an Associate Professor of Information Sciences at the University of Pittsburgh. His primary research areas are networked information systems and information fusion with emphasis on large-scale heterogeneous data integration, as well as wireless and sensor data management. He has over 15 years of active research experience in those areas. He has published more than 70 papers and his research has been well funded including support from NSF, EU, and the Norwegian Research Council. His experience in reliable information fusion and integration involves a wide spectrum of research projects ranging from stream and sensor data management and automatic information fusion in multi-robot search and rescue mission, to large-scale historical data integration from distributed Internet repositories.

GRANT11484019 -- Attachments-ATT6-1239-biographies.pdf

World-Historical Gazetteer
Data Management Plan

This Level I Start-Up proposal is primarily intended for working-group discussion rather than for data development. However, we will also build a testbed environment based on the 1.4 million georeferenced entities in English Wikipedia, potentially supplemented by some thousands of entities from other sources. The testbed is the data that we are describing in this plan.

The gazetteer testbed will be integrated with the existing Archive of the NSF-funded Collaborative for Historical Information and Analysis, which already exists in preliminary form. All decisions on gazetteer content and metadata management will be reviewed during the 2014 workshop, and the resulting plan will be implemented thereafter. As a result, this Data Management Plan includes elements of both the current set of actual data management, and the issues to be decided for the gazetteer to be created through this project.

Existing Data

Three categories of existing data stand out at present. 1) Data of the World-Historical Dataverse archive (www.dataverse.pitt.edu), housed at the Dataverse Network at Harvard University. 2) Data now incorporated into the CHIA Archive through the Col*Fusion procedure. These include the "Tycho" dataset on disease surveillance in the U.S., the Colonial Legacies dataset developed by John Gerring of Boston University, and the "20th-century global prototype" dataset developed by Patrick Manning and Chelsea Mafrica. 3) Other existing data for contemporary times, notably WikiData and GeoNames.

Gazetteer Data and Gazetteer Development

The objective of this project is creation of a CHIA gazetteer system, which will be permanent and may be either centralized or distributed. The gazetteer data are expected to be permanent, as they are necessary for documentation of the content in the historical archive. Periodic updates of gazetteer data will result as knowledge advances.

The NEH Digital Start-up project itself, in contrast, is a testbed system. This is the project-level work of developing standards for the permanent system. It may evolve into something different as an element of a persistent, long-term, stable CHIA architecture.

Archiving, Storage and Backup

All of the data stored in the CHIA archive will benefit from a program of Digital Stewardship to ensure preservation, documentation, careful identification of sources, and easy access to the materials for public users. Two repositories will hold the data for the gazetteer. First is a server located in and administered by the School of Information Science at the University of Pittsburgh, which is dedicated to the Collaborative for Historical Information and Analysis. Second, to the degree that the volume of data and the intensity of calculations require, data will be stored within the facilities of the Pittsburgh Supercomputer Center, with which CHIA has a contract through XSEDE. Backup procedures in each of these institutions are according to the standard procedures for each.

Responsibility

Dr. Kai Cao, postdoctoral associate at the World History Center, has responsibility for administering the gazetteer data stored and utilized.

Security

For all computers within the firewall of the University of Pittsburgh, the office of Computer Services and Systems Development (CSSD) takes responsibility for maintaining security and monitoring all campus internet installations.

GRANT11484019 -- Attachments-ATT7-1240-data.pdf

Audience and Access

Users of CHIA and therefore of the gazetteer documenting its data are expected to include a wide community of researchers, teachers, and students. It will be accessible through the CHIA website (www.chia.pitt.edu).

Ethics

The issues of ethics in handling of the World-Historical Gazetteer data is on the agenda for the August 2014 workshop. The decisions adopted at that time will be implemented.

Intellectual Property

The five universities collaborating in the NSF-supported award to CHIA have signed a Memorandum of Understanding governing issues of intellectual property created through work of the collaborative. This MOA (completed in April 2013) can be extended to the participants in the World-Historical Gazetteer.

GRANT11484019 -- Attachments-ATT7-1240-data.pdf

Appendix A: "World-Historical Gazetteer" Workshop Agenda (August 2014)

The workshop will work through general discussion, mostly divided into 90-minute sessions separated by 30-minute breaks (the latter to facilitate conversation). Each discussion will be led by a convenor selected from the group of participants, and there will be an audio record of the proceedings. In advance of the workshop, position papers (prepared by the convenor of each session) will be circulated, accompanied by brief case studies as appropriate.

DAY 1.

8:30. Coffee and light snacks.

9:00 – 9:30. **Welcome.** Ruth Mostern, Pat Manning

Welcome, introductions, review of workshop objectives.

9:30 – 10:30. **Session 1.**

Where are we now?

Coverage and gaps of world regions since 1500. How do we leverage existing content; when do we start from scratch?

10:30 – 11:00. Break

11:00 – 12:30. **Session 2.**

What types of entity should be covered?

"Places," units or features? Links to issues about relationships with existing gazetteer data standards:

12:30 – 2:00. Lunch

2:00 – 3:30. **Session 3.**

Alignment with existing non-historical gazetteers?

How desirable is it to link with Geonames, Wikidata, Open Street Map?

3:30 – 4:00. Break

4:00 – 5:00. **Session 4.**

Data models for gazetteers?

Typology and documentation. Gazetteers for vaguely-documented places. How much support should there be for onomastic research?

5:00 – 6:00. Reception, provided by World History Center.

7:00 – 8:30. Dinner, provided by World History Center.

DAY 2

8:30. Coffee and light snacks.

9:00 – 10: 30. **Session 5.**

Centralized or distributed infrastructure?

Alternative models for short term and long term.

10:30 – 11:00. Break.

11:00 – 12:30. **Session 6.**

Assimilating existing gazetteers?

How best to get people to contribute. Practical problems of data ingest.

12:30 – 1:30. Lunch

1:30 – 3:00. **Session 7.**

Extending the gazetteer through crowd-sourcing?

Gathering individual entries and individual additional names; selection of sources.

3:00 – 3:30. Break

3:30 – 5:00. **Session 8.**

Research design for a world-historical gazetteer?

Organizing development work in principal areas of need; potential funding sources for support of the remaining tasks.

5:00. Adjourn.

GRANT11484019 -- Attachments-ATT9-1242-appendices.pdf

Appendix B: Related Initiatives

- CHIA, the Collaborative for Historical Information and Analysis (http://chia.pitt.edu), directed by Patrick Manning and funded by the National Science Foundation, is a multi-institutional collaborative of scholars in social, natural, and information sciences. It is the parent organization of the World-Historical Gazetteer initiative. It is structured as a Research Collaborative that links participating institutions which are collecting data on population, climate, and other topics with a crowdsourcing tool (Col*Fusion, see below) to demonstrate the feasibility of building a continuously growing collection of diverse historical data and metadata. CHIA aims to develop a repository to house the incoming data and permit global and interactive analysis, to collect and process historical data, broaden the community of social and natural science researchers, analyze historical patterns of global change, and share its resources with researchers, policy-makers, teachers and students. CHIA is headquartered at the University of Pittsburgh with participating research groups at Boston University, Harvard University, Michigan State University, and University of California-Merced. To understand global social patterns as they exist today, it is increasingly clear that we need to understand how they have evolved over recent centuries. The Center for Historical Information and Analysis responds to this need and takes historical analysis into the realm of Big Data. It is expected that the data resources will grow to several terabytes in size. This project will stimulate development of more efficient research collaborations, enabling systematic large-scale consolidation of diverse historical data sources. Once collected and integrated, the data repository and analytical system will allow scholars to address a wider set of questions testing hypotheses about long-term and short-term social change at the global scale and catalyzing an expansion of the evidence base in social sciences. For example, our understanding of important societal issues can advance by linking health to demography and by incorporating climate and health factors into economic studies. Disciplinary theory will advance through interaction among the various scientific fields, so that a global network of social-science researchers will emerge.
    - Col*Fusion. Col*Fusion is a two-step application that enables the collection, archiving, documentation, and integration of discrete data files for CHIA. The application is under development at the University of Pittsburgh under the leadership of Vladimir Zadorozhny. Its first step is that of a crowd-source interface enabling users to contribute files to the CHIA archive with an internet connection through an API. In the second step, after initial processing of ingested files, the main step of Col*Fusion locates and links similar variables across multiple files, effectively merging the files for common analysis. In addition, the application documents the merged files with a "target schema," a common set of metadata that is extended as new files are added to the archive. Such merging of files has previously been possible only when their data-description formats were identical, *ex ante*.
- Pelagios (http://pelagios-project.blogspot.com/) is a project, under the direction of Leif Isaksen, Pau de Soto, Rainer Simon, and Elton Barker, to bring linked geographical data to the study of the ancient world. Its partners notably include Pleiades, directed by Tom Elliott, which developed the first significant ancient world gazetteer (http://pleiades.stoa.org/home) beginning in 2005. Pelagios 3 has recently been funded by the Mellon Foundation to annotate, link and index place references in digitized Early Geospatial Documents, documents that use written or visual representation to describe geographic space prior to 1492. The project has three primary objectives: (i) to provide an index of toponyms attested, and the places they refer to (where known), in all available EGDs, accessible both as Linked Open Data – i.e. freely downloadable content recorded in the Resource Description Framework (RDF) format – and via the Pelagios Web Service; (ii) to create an open and semi-automated toolset that allows the scholarly community to enhance and refine the index incrementally, by annotating for themselves place references in further historical sources (written and visual) as and when they are digitized; (iii) to develop a freely available analysis workbench and contextualization widgets that will enable researchers to bring together spatial documents in new and innovative ways and provide key contextual information as embedded content in third-party websites.

GRANT11484019 -- Attachments-ATT9-1242-appendices.pdf

- Regional Historical Gazetteers. Regional historical gazetteers of widely varying temporal and geographical scope and technical format have been under development for almost 20 years. Among them, those of the greatest scope and technical sophistication include the following:
    - Great Britain Historical GIS (http://www.port.ac.uk/research/gbhgis/), Vision of Britain (http://www.visionofbritain.org.uk/), and PastPlace. These three interlinked projects have been developed by Humphrey Southall at the University of Portsmouth. The Great Britain Historical Geographical Information System is a unique digital collection of information about Britain's localities as they have changed over time. Information comes from census reports, historical gazetteers, travelers' tales and historic maps. A separate website, created by funding from the UK National Lottery and extended and re-launched with funding from the Joint Information Systems Committee, makes this resource available on-line to everyone, presenting information graphically and cartographically. That site is called *A Vision of Britain through Time* and presents the history of Great Britain through places. Past Place is a new initiative to build a Web-based extensible Historical Place Index which depicts not just where places are, but directly links them to the old maps in which they are attested, and enables them to be searched for by the multitude of names that have been used for them over the centuries. This will transform access to historical maps for those interested in the history of a locality, and open the door to further integration with online content such as censuses, church and county records and local, national and international resources associated with them. It will do so by using public-friendly interfaces, Linked Open Data and webmapping technologies that connect global gazetteer entries to online images from historical maps drawn from online collections. For the world as a whole it will build on an existing URI-based digital gazetteer, making it easy to link to and from other major online resources, such as Wikipedia, so providing them with historical depth.
    - China Historical GIS (http://fas.harvard.edu/~chgis). This is a joint project of Harvard University and Fudan University (China). At Harvard, Peter Bol is the principal investigator and Merrick Lex Berman is the project manager. The China Historical Geographic Information System, **CHGIS**, was launched in January 2001 to establish a database of populated places and historical administrative units for the period of Chinese history between 221 BCE and 1911 CE. CHGIS provides a base GIS platform for researchers to use in spatial analysis, temporal statistical modeling, and representation of selected historical units as digital maps. The main objective of the CHGIS project is to create a flexible tool, in the form of a documented database of places and adminstrative units, which can be used to investigate any sort of geographically specific data related to China. The unique ID numbers for each of the CHGIS temporal instance records can be used as geocodes in relational databases, or to mark up texts, enabling users to import their own datasets into the CHGIS platform. Considerable flexibility has been designed into the GIS, allowing for alternate versions and variations of feature attributes, spatial data, and competing political entities. The CHGIS aims to build a reliable database of administrative units and settlements, but does not wish to impose a closed interpretation on the relationships among those units.
    - National Historical GIS (https://www.nhgis.org/). The National Historical Geographic Information System (NHGIS), developed by Steven Ruggles and collaborators at the Minnesota Population Project, provides aggregate census data and GIS-compatible boundary files for the United States between 1790 and 2011. For the periods prior to 1950, the highlights of the NHGIS include state and county aggregate data and boundary files (1790-2011), and census tract aggregate data and boundary files (1910-2011).
- Contemporary Global Gazetteers. Two global gazetteers of the contemporary world are notable for several characteristics: the number of entities they contain, their successful adoption of crowd-sourced method, and their use of linked-data methodologies that makes them amenable to linkage with other gazetteer projects with RDF-format content.

GRANT11484019 -- Attachments-ATT9-1242-appendices.pdf

- GeoNames (http://www.geonames.org/) . The GeoNames geographical database, founded by independent developer Marc Wick, is available for download free of charge under a creative commons attribution license. It contains over 10 million geographical names and consists of over 8 million unique features including 2.8 million populated places and 5.5 million alternate names. All features are categorized into one out of nine feature classes and further subcategorized into one out of 645 feature codes. The data is accessible free of charge through a number of web services and a daily database export. GeoNames is already serving up to over 30 million web service requests per day. It features a highly useable API.
- Wikidata (http://wikidata.org) is a new project of the Wikimedia Foundation: a free, collaborative, multilingual, secondary database, collecting structured data to provide support for Wikipedia, Wikimedia Commons and the other Wikimedia projects. Unlike Wikipedia, Wikidata will collect data in a structured form. This will allow easy reuse of that data by Wikimedia projects and third parties, and will enable computers to easily process and "understand" it. There are approximately 1.4 million georeferenced entities in Wikidata. Many are historical and many are characteristic of "places" rather than administrative features, and they will be resolving each name with a URI for linked data purposes. The Wikidata map interface is online at https://dl.dropboxusercontent.com/u/172199972/map/index.html.

- *Placing Names: Enriching and Integrating Gazetteers* (Indiana University Press, in contract). This collection of essays about historical gazetteers is edited by Merrick Lex Berman, Ruth Mostern, and Humphrey Southall. It includes almost 20 articles by authors with a range of backgrounds, who describe the content of various regional and thematic historical gazetteers, the technical and semantic challenges in historical gazetteer development and integration, and the types of spatial and historical analysis that historical gazetteers facilitate.
- Library of Congress Toponym Conflation Tool. The Library of Congress toponym conflation tool (at gazetteer.in), developed by Shekhar Krishnan and Schuyler Erle, is a free and open-source tool that resolves multiple names for the same entity (e.g. Chung-kuo returns People's Republic of China and Taiwan), and provides a reference map, a stable URI, a GeoJSON code, links to multiple reference gazetteers, and additional information including a revision history, timeframe, related entities, and alternate names for each entity.

GRANT11484019 -- Attachments-ATT9-1242-appendices.pdf

Appendix C: Bibliography

Alexandria Digital Library, *Guide to the ADL Gazetteer Content Standard*, version 3.2, http://www.alexandria.ucsb.edu/gazetteer/ContentStandard/version3.2/GCS3.2-guide.htm.

Aucott, P., Von Luenen, A. and Southall, H. (2009) Exposing the history of Europe: the creation of a structure to enable time-spatial searching of historical resources within a European framework. *OCLC Systems & Services*, 25 (4). pp. 270-286.

Buckland, M., and L. Lancaster, 'Combining time, place and topic', *D-Lib Magazine* 10 (2004).

Buckland, M., A. Chen, F. Gey, R. Larson, R. Mostern and V.Petras, 'Geographic search: catalogues, gazetteers and maps', *College and Research Libraries,* 68 (2007), 376-387.

Crane, G., and D. Smith, 'Disambiguating geographic names in a historical digital library', *ECDL '01 Proceedings of the 5th European Conference on Research and Advanced Technology for Digital Libraries* (2001) (perseus.mpiwg-berlin.mpg.de/Articles/geodl01.pdf, last accessed 5 Sept 2011).

Curry, M. R.,'Toward a geography of a world without maps: lessons from Ptolemy and postal codes, *Annals of the Association of American Geographers*, 95 (2005), 680–691.

Delano-Smith, C., 'Milieus of mobility: itineraries, route maps and road maps', in J. Akerman, ed., *Cartographies of travel and navigation* (Chicago, 2006), 16-68.

Goodchild, M. and L.L. Hill, 'Introduction to digital gazetteer research', *International Journal of Geographical Information Science*, 22 (2008), 1039-1044.

Goodwin, J., C. Dolbear and G. Hart, 'Geographical Linked Data: the administrative geography of Great Britain on the semantic web', *Transactions in GIS*, 12, suppl. 1 (2008), 19-30, esp. 26-7.

Hahmann, S. and Burghardt, D. 2013. How much information is geospatially referenced? Networks and cognition. *International Journal of Geographical Information Science* 27:1171-89.

Hill, L., 'Core elements of digital gazetteers: placenames, categories and footprints', J. Borbinha and T. Baker, eds., *Research and Advanced Technology for Digital Libraries* (Berlin, 2000), 280-90.

Janowicz, K. (2009). The role of place for the spatial referencing of heritage data. In *The Cultural Heritage of Historic European Cities and Public Participatory GIS Workshop* (pp. 17-18).

Jones, Christopher B., Harith Alani, and Douglas Tudhope. "Geographical information retrieval with ontologies of place." *Spatial information theory*. Springer Berlin Heidelberg, 2001. 322-335.

Mostern, R. 'Historical gazetteers, an experiential perspective, with examples from Chinese history', *Historical Methods,* 41 (2008), 39-46.

Mostern, R. and I. Johnson, 'From named place to naming event: creating gazetteers for history', *International Journal of Geographical Information Science*, 22 (2008), 1091-1108.

Petras, V., R. Larson and M. Buckland, 'Time period directories: a metadata infrastructure for placing events in temporal and geographical context', *JCDL '06: Proceedings of the 6th ACM/IEEE-CS joint conference on Digital libraries* (Chapel Hill, 2006), 151-160.

Scheider, S., and K. Janowicz. "Places as media of containment." *Proceedings of the 6th International Conference on Geographic Information Science (extended abstract, forthcoming)*. 2010.

Shaw, R., 'Event gazetteers for navigating humanities resources', *PIKM08: Proceeding of the 2nd PhD workshop on Information and knowledge management* (Napa Valley, 2008), 89-92.

Simon, Rainer, Joachim Korb, Christian Sadilek, and Rainer Schmidt. "Collaborative map annotation in the context of historical GIS." In *E-Science Workshops, 2009 5th IEEE International Conference on*, pp. 139-142. IEEE, 2009.

Smart, P.D., C.B. Jones and F.A. Twaroch 'Multi-source toponym data integration and mediation for a meta-gazetteer service', Geographic Information Science, *Lecture Notes in Computer Science*, 6292 (2010), 234-248.

Southall, H.R. 2003. Defining and identifying the roles of geographic references within text: examples from the Great Britain historical GIS project. In *HLT-NAACL 2003 Workshop: Analysis of Geographic References*, 69-78. Edmonton: Association for Computational Linguistics.

GRANT11484019 -- Attachments-ATT9-1242-appendices.pdf

Southall, H.R. 2012. Rebuilding the Great Britain Historical GIS, part 2: a geo-spatial ontology of administrative units. *Historical Methods: A Journal of Quantitative and Interdisciplinary History* 45:119-134.

Southall, H.R., Mostern, R., and Berman, M. 2011. On historical gazetteers. *International Journal of Humanities and Arts Computing* 5:127-45.

Tuan, Y.F., 'Place: an experiential perspective', *Geographical Review,* 65 (1975), 151-165.

GRANT11484019 -- Attachments-ATT9-1242-appendices.pdf