Proposal:

A Global Historical GIS (GH-GIS) Project

PIs:

**John Gerring**
Professor  
Department of Political Science  
Boston University  
232 Bay State Road  
Boston MA 02215  
jgerring@bu.edu

**Nathan Nunn**
Assistant Professor  
Department of Economics  
Harvard University  
Littauer Center  
1805 Cambridge Street  
Cambridge, MA 02138  
nnunn@fas.harvard.edu

**Robert Woodberry**
Assistant Professor  
Department of Sociology  
University of Texas, Austin  
1 University Station A1700  
Austin, TX 78712-0118  
bobwood@austin.utexas.edu

Collaborators:

**Peter Bol**
Professor  
East Asian Languages & Civilizations  
Director  
Center for Geographic Analysis  
Harvard University  
pkbol@fas.harvard.edu

**Wendy Guan**
Director  
GIS Research Services  
Center for Geographic Analysis  
Harvard University  
wguan@cga.harvard.edu

**Ben Lewis**
Senior GIS Specialist  
Center for Geographic Analysis  
Harvard University  
blewis@cga.harvard.edu

*Draft: March 15, 2009*
Every challenge facing the world today has a history -- in most cases, a very long one. These histories matter, not simply because we wish to understand the causes of a current problem but also because we wish to understand its trajectory. One’s perspective on an issue is necessarily affected by one knowledge of its past, or pasts. For example, one is apt to consider “globalization” quite differently if the topic is viewed in the context of (a) current history (the past half-century), (b) modern history (the past two centuries), or (c) human history (since the evolution of the human species).

We do not mean to imply that the longue durée always offers a superior vantage point. Sometimes, a present-ist perspective may be justified. Our argument is that the historical perspective(s) should always be taken into consideration. Sometimes, there is valuable information to be gained by examining the long-run evolution of a topic or problem, and one cannot judge the issue until such an accounting has been completed. We suspect that for most topics of global interest -- including poverty, disease, premature mortality, social conflict, autocracy, poor governance, environmental degradation, and globalization – a strong case can be made for the importance of history.

Unfortunately, many topics suffer from a dearth of readily accessible and globally comparable historical data. This scarcity is partly a product of disciplinary proclivities. Historians generally prefer to approach subjects with a microscopic lens. Historical knowledge therefore tends to be highly compartmentalized – by region, by era, and by subject. Unfortunately, these detailed narratives do not cumulate readily into a coherent global account. Historical narratives are not, as a rule, neatly commensurable.

By a “coherent” or “unified” historical account we do not mean to imply that such an account of the world must be monocausal, uniform, or linear. One may find a history of disjunctures and causal interactions; X may affect Y differently in different times and places. The point is, this sort of conclusion requires comparing X and Y systematically across regions and time-periods. It requires a global accounting of historical processes. And this, in turn, requires a considerable reduction of the historical record -- into units that can be directly compared through time and space, i.e., data.

A few social scientists have taken up this challenge, presenting history on broad tableaux. One thinks of the pathbreaking work of Paul Bairoch, Fernand Braudel, Jared Diamond, Eric Jones, William McNeill, Douglass North, Immanuel Wallerstein, and the research teams of Engerman/Sokoloff and Acemoglu/Johnson/Robinson, as well as countless books written in a more popular vein.1 Studies in this genre illustrate the promise, and also the pitfalls, of doing history on a grand scale. Each author (or team) must construct a dataset more or less from scratch, or resign herself to sweeping generalizations without strong evidentiary support. Scholars are often able to measure inputs (inferred from the present properties of the world) and outcomes (measured in the contemporary era) with a fair degree of precision. However, what lies in between – the meat and gristle of the argument – frequently remains mysterious. Mechanisms are difficult to trace.

---

Similarly, those who teach history (at secondary or university levels) must put together materials from a wide range of disparate sources, often in different languages and various data formats. It is a painstaking endeavor. Alternatively, teachers may rely on textbooks; but this generally precludes, or at least impedes, the inclusion of original course content. Policymakers, amateur enthusiasts, and politicians who dabble in history are in the same unfortunate position. They are compelled to spend years constructing an original dataset, or must jettison their global ambitions for a much smaller slice of reality (presumably more tractable, given time and resource constraints).

This is not a happy state of affairs. It has contributed to a pejorative view of history, which is often regarded as (a) essentially unknowable, (b) an idiographic story about who did what to whom (hence, without relevance for the present), or (c) a rich mine from which virtually any narrative might be spun (hence, useful for mythmakers and prognosticators). Indeed, on those occasions when History has been brought to bear on questions of public deliberation, we suspect it has rarely been to good effect.

The partial remedy that we propose rests upon the application of a well-established technology to a relatively new subject, on a scale never previously attempted. For decades, geographic information system (GIS) techniques have been applied to problems in environmental and land management, military logistics, and geography. Recently, scholars have begun to appreciate its potential for historical analysis. In this quest, researchers have digitized maps of the past (either contemporary maps of historical subjects or very old maps), and – in a very limited way – incorporated features within these maps as attributes (“data”). Such projects typically cover only a few countries, situated mostly in the developed world (e.g., in Germany, Great Britain, and the United States). China is the one country in the developing world to have generated a substantial historical GIS project (see Appendix B).

The time is ripe for a truly global, historical GIS project, one that integrates extant GIS information with newly digitized data to make history accessible in a way that professional historians, social scientists, policymakers, and lay citizens can learn from and build upon. GIS techniques are uniquely suited for this task, allowing for the storage and retrieval of any quantity of data, along any number of parameters, for any number of (possibly overlapping) spatial units, and for any length of time. This flexible technology also provides multiple methods for visual presentation and for data analysis, as previewed below.

Of course, we do not mean to suggest that the mysteries of the past will be unraveled merely by collecting more data and presenting it in novel formats. GIS, by itself, will not solve recalcitrant problems of causal attribution. However, we do suppose that problems of evidence loom large in ongoing controversies about the past, and these are often directly relevant to problems faced in the present. GIS technology thus has the capacity to enhance the relevance of history for challenges facing the world today.

This proposal outlines an approach for creating a global historical GIS (GH-GIS) dataset, one that incorporates a wide range of spatially based data relevant to present concerns. The sections that follow lay out (1) basic features, (2) derivative features, (3) sources, (4) dataset construction, (5) infrastructure and applications, (6) conclusions, and (7) references. Appendices review the

---

qualifications of the research team (Appendix A), related web-based projects (Appendix B),
historical atlases (Appendix C), and additional data sources (Appendix D).

1. Basic Features

The aim of this project is to unify the world of spatial representation (GIS) with the world of
substantive information such that the two are interwoven seamlessly. That is, information about a
topic (e.g., passenger traffic on a railroad line) is joined with a shape file showing the location of that
topic (e.g., the location of the railroad line), and traced through time. Granted, not all topics lend
themselves to this sort of global representation. Some topics are rightly understood as idiographic
(there is no point or little possibility in standardizing the feature). However, the topics of greatest
concern to academics, policymakers, and the lay public are usually generalizable. Thus, we have
every reason to expect that the issues investigated in this project will include important, and relevant,
features of the past -- not flotsam and jetsam.

We begin with places, e.g., empires, countries, colonies, regions, cities, towns, and
administrative units. Names will be recorded in the language indigenous to that region, in English,
and in any other language in which data might be collected (e.g., French, German, and other
“world” languages). For each name, alternate spellings will be included.

The gazetteer portion of the project will consume a good deal of time and attention. Although
there exist several contemporary gazetteers to draw upon (notably, the NGA GEOnet Names
Server [GNS]), historical gazetteers are rare and generally cover only a single country or continent. A
project of this nature requires the construction of a more or less comprehensive global historical
gazetteer -- which, when completed, will be the first of its kind. Fortuitously, the gazetteers
developed for AfricaMap and the China Historical GIS at the Center for Geographic Analysis at
Harvard contain 25-30 million place names and alternate names. These will serve as our point of
departure.

For each place-name, we aim to identify a spatial location, which may vary through time.
Note that boundaries between political units are often difficult to identify (even at a particular point
in time), and are likely to change over time. “France” has not always referred to the same
cartographic entity. This makes the task of geo-referencing difficult, but also of supreme
importance.

For most spatial entities, additional historical data is likely to be available (other than mere
location). These additional data elements fall into eleven categories: (1) demographics, (2)
government, (3) global travel, (4) sovereignty, (5) conflict, (6) infrastructure, (7) commerce, (8)
education, (9) religion, (10) language, and (11) ethnicity. The basic unit of observation is thus both
spatial (e.g., a specific city or country) and temporal (generally, a year).

Demographic data includes (a) population, (b) population density, (c) mortality (per 1000), (d)
infant mortality (per 1000 live births), (e) life expectancy, (f) births (per 1000 inhabitants), and (g)
fertility (births per woman). Each of these statistics will be collected for various spatial units (e.g.,
empires, countries, colonies, cities,...) and for various subgroups, as available (e.g., male/female,
ethnic groups, religious groups, missionaries, racial groups, settlers, indigenes, citizens/noncitizens,
slave/free).
**Government** data include (a) number of administrators or personnel, (b) revenue, and (c) expenditure. (At a later date, we might collect information on the names and ethnic/racial/linguistic/gender identities of top leaders.) Data will be collected for various political units (e.g., imperial, national, subnational, municipal,...) as well as for various departments and subdivisions (e.g., military, non-military, police, treasury, postal, transport,...).

**Global travel** data include (a) the year of first contact (the year in which a spatial unit is “discovered” by another unit), (b) the path of early explorers (traced through time), and (c) migration flows. The latter will be counted as the number of migrants moving from one area to another in a given period, thus taking account the point of origin and point of arrival (in a matrix format, on the model of bilateral trade data).

**Sovereignty** data include (a) legal sovereignty of a spatial unit (the international recognition of a unit’s right to self-determination), (b) effective sovereignty (a more complex coding of effective self-determination [Adams, in process]), (c) colony (a territory without internationally recognized sovereignty that belongs to another political unit to which it is not fully integrated, aka dominion or overseas territory), and (d) contest territory (territory whose sovereign status is actively contested, i.e., where no authority enjoys an effective monopoly of force).

**Conflict** data include (a) military attacks (understood as any violent engagement with multiple casualties, mapped to the specific geographic location where the casualties were inflicted and associated with the actual number of casualties, if known), (b) regions characterized by violent conflict (at some time during a given year).

**Infrastructure** data include (a) waterways, (b) roads, (c) railroads, (d) post, (e) telegraph, (f) telephone, and (g) electricity. For each infrastructural feature, we would take note of its spatial location (e.g., the location of waterways, roads, railroads, postal bureau, telegraph lines, and telephone trunk lines), various measures of quality, wherever possible (e.g., paved/unpaved roads, deep/shallow waterways), and measures of volume (traffic, as measured by boat passages, freight, passengers, mail items, telegrams,...). The foregoing data could also be aggregated by spatial unit (empire, nation, region, city, town).

**Commerce** data includes trade, as measured by (a) total imports, (b) total exports, (c) total bilateral trade flows, and (d) trade in various specified commodities. Commercial statistics also include the location of (e) trading posts, (f) mines (classified according to mineral), (g) crops (classified according to commodity group), and (h) arable land.

**Education** data includes (a) education expenditures, (b) enrollment, (c) male/female enrollment ratio, and (d) location of schools. These attributes will be noted for each spatial unit (e.g., country, territory, city) as well as for different types of educational systems (e.g., missionary, secular private, public) and education levels (primary, secondary, university) within a spatial unit.

**Religion** data includes, for each religion or sect, (a) the area in which a religion is regularly practiced (by more than a few adherents), (b) the area in which it is the majority or dominant faith, (c) the total number of adherents (within a given jurisdiction), (d) density (percent of population who adhere to a religion), (e) places of worship (location), and (f) schools (location, and number of students).
Language data includes, for each language, (a) the area in which it is regularly practiced (by more than a few adherents), (b) the area in which it is the majority or dominant language, (c) the total number of speakers (within various jurisdictions), (d) the total number for whom it is a first language (mother tongue), and (e) density (percent of population who speak it as a first or second language).

Ethnicity data includes, for each ethnicity, (a) the area in which an ethnic group is prevalent (a majority or significant minority), (b) the area in which it is the majority or dominant ethnic group, (c) the total number of group members, and (d) density (percent of population who are members of that group).³

These eleven categories, with their various sub-components, comprise the primary variables in the dataset. We anticipate that additional primary variables will be added to this list as the project proceeds, and that distinctions noted above will become somewhat more differentiated.

2. Derivative features

To complement the primary data described above, a “secondary” dataset will be prepared at the end of the project in order to present the material in an accessible fashion for non-specialists. Here, we will include only one data source for each variable. That is, we will make judgments about the veracity of multiple data sources, presenting only that source deemed most authoritative and comprehensive. Alternatively, if no single source is superior, a variable will be constructed from several sources (by imputation). In addition, missing data will be imputed wherever such procedures seem justified (i.e., wherever sufficient variables exist that might help predict the value of missing parameters). Note that the purpose of completeness in this context is to provide a dataset that can be used for varied analyses -- descriptive, causal, and predictive -- without biasing results by over-representing those parts of the world, or those eras, where richer data is available. Imputation techniques involve interpolating missing data for a given spatial unit through time between recorded data points and extrapolating missing data across spatial units. Estimates of uncertainty will be included, based on whatever imputation procedures are employed.

Using variables drawn from the secondary dataset, an additional set of variables may be generated (either by the authors of this project or by end-users). For example, data collected for Protestants and Catholics may be combined into the larger category “Christian,” and from thence combined with data collected for Muslims, Jews, Buddhists, Hindus and other religions to form the superordinate category “religion” – or, at an even higher level of aggregation, “ascriptive identities.” Note that basic data is collected at the most disaggregated level possible in order to allow for the greatest possible flexibility; data can almost always be aggregated up, but can never be disaggregated down.

Another sort of derivative variable involves a ratio comprised of several basic variables. Note that many of the basic variables are best understood as a share of total population, of total trade, or of total land area (calculable with GIS software) – rather than in a “raw” format. More complex

³ For present purposes, ethnicity is defined residually – as any important group identity (as understood by historical subjects) not isomorphic with religion or language. This precludes the problem of double-counting some group identities (e.g., Judaism), while allowing researchers to re-code the category to suit their interests and historical judgments.
variables may also be created. For example, one may calculate relative homogeneity/heterogeneity by examining the relative shares of a property (e.g., religion, ethnicity, language, race) within a population group or land area. (Heterogeneity formulas vary, depending upon one’s theoretical interest.) Distances between spatial entities may be calculated with GIS software, using various algorithms (e.g., shortest distance between two points, shortest travel distance, and total travel-time given contemporary transport facilities). This, in turn, allows for complex “network” analyses.

Several highly abstract concepts of immense historical and contemporary importance deserve special mention. These include a) economic development, b) political development, c) legitimacy, d) cultural change, e) international power, and f) colonialism/hegemony.

Prior to the arrival of wage-based economies it is difficult to identify summary measures of economic development. “GDP” is a problematic concept in the eighteenth century. We have the sense that parts of Western Europe and East Asia were more advanced than sub-Saharan Africa, but we have difficulty operationalizing this intuition. Here, trade data, coupled with demographic indicators such as urbanization and the population of cities, is helpful. Since the formal sector of most economies were linked to long-distance trade, the existence of import and/or export data is a strong indication of the relative development of all but the largest and most insular economies (e.g., Manning 1982: 4). Moreover, where territories were urbanized and/or where the “primate” city was populous we can assume that levels of wealth, technology, and social and political organization were high, relative to other territories or other time-periods (Acemoglu et al. 2002; Bairoch 1988). Fortuitously, most civilizations that we are aware of – and certainly all modern civilizations -- were based on urban agglomerations. In the twentieth century, and perhaps even the nineteenth, it becomes possible to arrive at reasonably good estimates of GDP per capita (Maddison 2001), which can be combined with trade and demographic data from earlier periods to arrive at a comprehensive accounting of economic development throughout various historical eras.

Political development in a state may be measured across several dimensions: the size of the bureaucracy (as a share of the total population), revenue (as a share of total output, exports, or population), and expenditures (as a share of total output, exports, or population).

As a measure of state legitimacy, we suggest counting the size of the domestic security force (which may consist of the military, the police, or a separate force mobilized for the purpose of keeping internal order) as a share of a) the general population and b) of the total land area. Presumably, legitimate states are able to sustain order with smaller investments in domestic security.

Constructing accurate and sensitive measures of cultural change is a daunting task. One set of proxies involves linguistic and religious practices. If these change – if, for example, a region adopts a new language and/or religion – there is strong reason to suppose that a wide-ranging cultural transformation is also occurring. The speed and thoroughness of this transformation can presumably be tracked by the rate and extent to which indigenous practices disappear. Thus, variables measuring linguistic and religious practices offer a crude indicator of broader societal

---

4 Bairoch (1988), Chandler (1987), Childe (1950), Modelski (2000). Partial exception might be made for the Egyptian empire (Modelski 2000: 25-6) and the Roman Empire in its later stages, when wealth migrated from Rome to large latifundia-style estates situated in rural regions around the empire. However, the fact that this movement was associated with the empire’s decline is not coincidental.
transformations. A related approach looks to changes in the racial complexion of a population as a
due to the cultural transformation of societies and – equally important -- how integrated/segregated
these societies were, overall. The number and composition of schools is a final measure of cultural
change, given the presumed importance of formal education in establishing and maintaining social
identities in the modern era.

The international power of a state may be measured by a) its military (permanent military
personnel and military expenditures), b) its land area (the area over which it exercises sovereignty),
and c) its population (the number of people over whom it exercises sovereignty). Granted, it is
sometimes tricky to determine questions of sovereignty (who controls which peoples within which
territory).

With respect to colonialism, one might wish to recognize two fundamental dimensions: a) duration
and b) intensiveness. The former may be measured in years (although it should be noted
that dating the onset of colonial relationships is sometimes complicated, even in the case of overseas
relationships). The latter is much more complex, both conceptually and operationally, including (at
the very least) a) the number of staff from the metropole allocated to the colony (as a share of the
colonial administration and as a share of total population in the colony), b) the number of military
personnel from the metropole (as a share of the colonial administration and as a share of total
population in the colony), c) the number of settlers from the metropole (as share of total population
in the colony), d) legal penetration (the number of customary court cases/total number of court
cases), e) expenditures by the metropole for the upkeep of a colony (as a share of total exports or
total population in the colony).

Many other examples could be pursued. The general point is that the basic-level data
categories contained in Section 1 are highly flexible, and may be aggregated in multiple ways so as to
address larger issues of historical and contemporary concern.

3. Sources

Data sources for a project of this magnitude are, of necessity, varied. Here, we identify major
categories of primary and secondary sources. For a more detailed accounting, the reader is referred to
appendices B, C, and D.

First, we will work to integrate extant GIS projects that speak to long-term historical
developments. This includes: Colonialism and its Legacies (Boston University, Northwestern
University), the China Historical GIS Project (Harvard University), Digital Atlas of Ancient
Civilizations in Europe and Middle East (Harvard University), Euroatlas (Switzerland), AfricaMap
(Harvard University), the Project on Religion and Economic Change (UT Austin), the
Animated Atlas of African History (Brown University), the Electronic Cultural Atlas Initiative (UC
Berkeley), the Mapping History Project (University of Oregon, University of Munster), the
Centennia Historical Atlas (cd-ROM), the World Language Mapping System (Colorado Springs,
CO), and various European and North American country-level GIS projects (in the US, the UK, and
elsewhere). See Appendix B. Note that where data are proprietary, integrating it into the present
project will involve negotiating with the copyright holders; special compensation may be required.
Second, we will scan and geo-reference additional cartographical information, drawn from historical atlases and antique maps. The purpose of these maps is (a) to clarify the precise location of spatial units of importance to the project, especially those that undergo changes through time (e.g., border changes); and (b) to add new data elements to the database (maps are a primary data source of considerable importance). Special attention will be paid to areas of the world neglected by prior h-GIS projects so that a truly global representation of history can be created.

A selection of important historical atlases is listed in Appendix C. Old maps are available from a number of sources including the David Rumsey Map Collection, the Electronic Cultural Atlas Initiative (ECAI), Gallica, the Mapping History Project, AfricaMap, Wikipedia, and Query and Context based Visualization of Time-spatial Cultural Dynamics (QVIZ). The holdings of the Harvard Library system are immense, and there are extensive map collections in many libraries around the world. (Typically, libraries will scan a map for a small fee and send the pdf by zip file or by mail as a compact disk. Thus, it is not necessary to traverse the world in order to gather information from maps so long as a collection is properly catalogued.)

Third, we will work to integrate useful non-cartographic data sources into the GIS format, so that the data (e.g., expenditures by a local government) are linked (as “attributes”) to the correct geographic entities (i.e., that particular municipality, with its correct coordinates). This part of the project builds on extant historical data collection efforts such as the Colonial Legacies project, as well as on a vast range of primary and secondary materials – a preliminary list of which is included in Appendix D.

4. Dataset construction

Each observation in the primary dataset will note the following: (a) the common name of the entity, if any (including alternate names and alternate languages), (b) the location of the entity (vector or raster data in GIS format), (c) attributes of interest (substantive information, as in Section 1), (d) year (and more specific date, if any), (e) source (which may involve several layers of sources), and (f) additional notes. The latter is an all-purpose field allowing us to comment on the viability of the source, disagreements among sources or coders, special coding rules, or any other facet of the data that might be relevant to future users.

This information system should make the task of any future re-coding immeasurably easier and allows for a full reporting of the procedures employed. Replicability must be maintained, at all costs. Users of the dataset should be able to reconstruct the origin of each piece of data.

It should be emphasized that many of the coding categories listed in Section 1 involve some interpretation (by virtue of meaning different things in different contexts). For example, project leaders will need to decide how to define “waterways,” “arable land,” “ethnicity,” and all the hotly contested political variables (sovereignty, colony, et al). In some circumstances, multiple operationalizations may be employed (the same general concept measured in alternate ways).

The categorization scheme in Section 1 is indeed somewhat arbitrary, representing only one possible way of dividing up the material. Some categories overlap (e.g., data on missionary schools is evidence of both education and religion). Even so, there is a need to organize the material in some fashion. For purposes of data collection and organization, an ontology (hierarchical taxonomy) is
essential. End-users may create their own *folksonomies* (informal classification schemes), depending upon their interests.

Note that the purpose of this project is to capture as much of the “feel” of the original historical context as possible, while also allowing for systematic comparisons across time and space. This means that original meanings must be retained, but that these meanings must often be re-coded so as to render them commensurable with similar information drawn from different eras and areas. Thus, a territory referred to by contemporaries as a “protectorate” may be relabeled a “colony” in a more general system of classification. The categories listed in Section 1 represent an attempt to create universally applicable concepts which re-frame – but do not replace -- those with local resonance. We recognize that some “conceptual stretching” is involved here (Collier, Mahon 1993; Sartori 1970). However, we comfort ourselves that end-users may always return to the original terms by which an entity was understood, or provide their own superordinate categories.

Another source of error is more basic, stemming from the nature of the evidence at hand. Information necessary to code the numerous topics identified in Sections 1 and 2 is not ubiquitous - - and, where available, may not be terribly reliable or precise. In general, we anticipate that the quantity and quality of data will be highest for Europe, Japan, and China, and will diminish everywhere as one moves back in time. At any given point in time some factors (e.g., population) will be easier to estimate than others (e.g., fertility), and a few dimensions will be impossible even to guess.

Several steps will be taken to deal with error, uncertainty, and incompleteness. First, PDF copies of all original data sources will be preserved, and (if possible) linked directly to the dataset so that users may reference the original whenever issues of interpretation or fact-checking arise. Second, all data will be collected and retained in a format that hews closely to the context of the primary source, as indicated (“primary” data). Third, multiple data sources will be employed wherever available, and especially wherever we have reason to doubt the veracity of a source. Finally, and perhaps most importantly, we anticipate that an ongoing audit for faulty or incomplete data will be provided by the community of end-users. In order to facilitate this, a “wiki” format will be adopted that allows anyone to comment on particular features of the data, or provide additional data. Managerial tasks (to incorporate these amendments and emendations) will be divided among specialists in different areas.

5. Infrastructure and Applications

In developing the architecture for GH-GIS, ease of use will be a primary consideration. The resulting dataset must be accessible to non-technical people so that the range of potential users extends beyond the small ambit of GIS-savvy experts. To that end, researchers and members of the lay public will be consulted frequently to guide the ongoing design of the user interface.

The architecture of GH-GIS will provide a system for storing, querying, displaying, and collaboratively editing historic spatial information. The system will be decentralized to enable multiple organizations to collaboratively build a variety of types of historical atlases at a variety of scales from regional to global. The design of the system will make it possible to store and access a virtually unlimited volume of mapped information. The portion of data stored at Harvard will be
made permanently available as it will be stored in systems maintained by permanent endowments such as the Harvard Geospatial Library (http://hul.harvard.edu/ois/systems/hgl/) and the DataVerse Network (http://thedata.org/).

The software infrastructure used in GH-GIS will build on the Harvard University Infrastructure for Geospatial Collaboration (HUG), recently developed for the AfricaMap project (http://AfricaMap.harvard.edu) by Suzanne Blier and Ben Lewis at the Center for Geographic Analysis at Harvard (CGA). The HUG infrastructure provides several essential pieces for a global historical GIS. This includes:

1) A web-based mapping client which supports the following user functions:
   a. Ability to turn layers on and off and control transparency, making it possible to view multiple layers concurrently.
   b. Rapid panning and zooming against multiple map layers.
   c. Ability to search the gazetteer by type and name, and quickly return matching features from millions of records.
   d. Ability to run text queries against multiple maps and highlight results.
   e. Ability to perform “drill down” queries against multiple map layers. A user clicks on a set of overlapping maps and returns information for the location clicked for all maps.
   f. Tight integration of Google, Microsoft, Yahoo or other commercial map providers with one’s own maps. For current public earth imagery, these systems are unmatched.
   g. Ability to view layers in Google Earth in addition to the web map client.
   h. Encapsulation in URL of the state of the web client including zoom extent, layers, display order, transparency settings, and text queries.
   i. Control of all layers in the system with access to information on Layer name, Description, Date, Source, Scale, Language, Legend, Zoom, reference, Google Earth view, Metadata access, download access.

2) A software mapping framework that supports the functions of the mapping client described above, and has the following characteristics:
   a. Ability to store very large sets of raster or vector data and make it accessible for query and display.
   b. Ability to cache map data to support very fast displays and high scalability.
   c. A service oriented architecture which makes its content available to other systems as web services and is also capable of consuming services which other organizations make available and want to share.
   d. Open source software which other organizations can build on and enhance.

3) Base maps that can be made available for use in GH-GIS. The AfricaMap project can contribute 18 layers composed of 3155 map sheets for the years 1612 to 1990, which can easily be extended.
Additional enhancements to HUG will be necessary to support the full requirements of a Global Historical GIS. This includes:

1) The development of a system for collaboratively building datasets derived from source materials. Such derived datasets may include historic layers such as trade routes, political boundaries, transportation routes, etc. This system will be built upon the successful model of Open Street Map (http://wiki.openstreetmap.org/) and will make use of the open source Open Street Map source code for building it. A key element in this system will be maintenance of provenance between source materials and derived features. As with a wiki framework, it will be possible to see who made what changes when, and to roll back changes if necessary.

2) The creation of a web accessible map layer index that can be accessed via an open API. This will support the centralized access and sharing of datasets and web map services between organizations, and will make it possible for other organizations to create its own web mapping client for their own purposes. (In the current AfricaMap application this index would populate the Layer List tab.)

3) Robust support for time range information (begin date, end date, level of certainty for begin and end dates) for every feature in the system.

4) Tools to support tagging of features in the system. A tagging ability would allow users to organize materials according to their own needs. When many people use tags they end up providing a dynamic and flexible alternative to predefined ontologies known as folksonomies.

6. Conclusion

The extraordinary flexibility afforded by GIS technology suggests that this project may become a basic resource – for teachers, scholars, policymakers, and citizens – in years and decades to come. Insofar as Google Earth serves as a foundation for geographical explorations of contemporary data, GH-GIS may serve as the principal entrée for historical data.

The wide range of substantive topics contained in the project assures its relevance for scholars across the social sciences and humanities -- including anthropology, archeology, demography, economics, education, geography, history, international relations, linguistics, political science, religion, and sociology. Moreover, the integrative nature of GH-GIS should facilitate collaboration across these disparate, and often fragmented, disciplines. Finally, we expect multiple uses of a more applied nature, whose shape and purpose can scarcely be anticipated in advance.

7. References


Appendix A:
The Project Team

John Gerring (PhD, University of California at Berkeley, 1993) is Professor of Political Science at Boston University, where he teaches courses on methodology and comparative politics. His books include Party Ideologies in America, 1828-1996 (1998), Social Science Methodology: A Critical Framework (2001; revised ed. planned for 2010), Case Study Research: Principles and Practices (2007), A Centripetal Theory of Democratic Governance (with Strom Thacker, 2008), Concepts and Method: Giovanni Sartori and His Legacy (ed. with David Collier; 2009), Global Justice: A Prioritarian Manifesto (in process), and Democracy and Development: A Historical Perspective (in process). His articles have appeared in a wide variety of political science journals. He served as a fellow of the School of Social Science at the Institute for Advanced Study (2002-03) and a member of The National Academy of Sciences’ Committee on the Evaluation of USAID Programs to Support the Development of Democracy (2006-07). He is the recipient of a grant from the National Science Foundation to investigate the global impact of colonialism (2007-10) and also serves as President of the American Political Science Association’s Organized Section on Qualitative and Multi-Method Research (2007-09).

Nathan Nunn (PhD, University of Toronto, 2005) is Assistant Professor of Economics at Harvard University, where he teaches courses in Economic Development and International Trade. He has published in a variety of journals including the Annual Review of Economics, Journal of Development Economics, and Quarterly Journal of Economics. He is a Faculty Research Fellow at the National Bureau of Economic Research (NBER), an Affiliate of the Bureau for Research and Economic Analysis of Development (BREAD), and an Associate Editor of the Journal of International Economics. In 2009, Professor Nunn was selected as an Alfred P. Sloan Research Fellow and fellowship recipient.

Robert Woodberry (PhD, University of North Carolina – Chapel Hill, 2004) is Assistant Professor of Sociology at the University of Texas at Austin where he teaches courses on statistics and on religion and global change. He is also director of the Project on Religion and Economic Change www.prec.com. He has published in a variety of journals including the American Sociological Review, Social Forces, and the Annual Review of Sociology and won the outstanding article award from the American Sociological Association (ASA), Sociology of Religion Section (2001). He has served on the council of ASA sections on sociology of religion and history of sociology and is the recipient of grants from the Templeton Foundation, Metanexus Institute, and Louisville Institute.

Peter K. Bol is a Harvard College Professor and the Charles H. Carswell Professor of East Asian Languages and Civilizations. He led Harvard’s university-wide effort to establish support for geospatial analysis in teaching and research; in 2005 he was named the first director of the Center for Geographic Analysis. He also directs the China Historical Geographic Information Systems project, a collaboration between Harvard and Fudan University in Shanghai to create a GIS for 2000 years of Chinese history, and is involved in other projects aimed at enhancing digital information linkages between East Asian and Western scholars.

Wendy Guan (Ph.D., University of Georgia, 1993) is Director of GIS Research Services at Harvard University's Center for Geographic Analysis. She has over 20 years of experience developing, applying, training, and managing geospatial technology in academic as well as government and private sectors. She taught GIS in various universities, including the Harvard Extension School.
**Ben Lewis** Before joining Harvard in 2007, Mr. Lewis was a project manager with Advanced Technology Solutions of Pennsylvania where he led the company in adopting platform independent approaches to Geographic Information Systems (GIS) development. Ben studied Chinese at the University of Wisconsin and has a Masters in Planning from the University of Pennsylvania. After Penn, Ben worked at the UC Berkeley GIS Lab, managed the GIS group for the transportation engineering firm McCormick Taylor, and coordinated the Land Acquisition Mapping System for the South Florida Water Management District. Ben is especially interested in technologies that lower barriers to GIS access.
Appendix B:
Related Projects

David Rumsey Map Collection
URL: www.davidrumsey.com/
Location: San Francisco, CA.
Sponsor: David Rumsey.
Principal: David Rumsey.
Period: 1680-1930.
Area: Mostly United States, but also a smattering of maps from elsewhere around the world.
Availability: A small portion of Rumsey’s collection is posted on-line.
Status: Ongoing.
Description: Maps are geo-referenced but no data is extracted.

[From the web site…] The over 120 historical maps in the Google Maps and Google Earth Rumsey Historical Maps sites have been selected by David Rumsey from his collection of more than 150,000 historical maps; in addition, there are a few maps from collections with which he collaborates. These maps can also be seen in the Gallery layer of Google Earth, Rumsey Historical Maps layer. All the maps contain rich information about the past and represent a sampling of time periods (1680 to 1930), scales, and cartographic art, resulting in visual history stories that only old maps can tell. Each map has been georeferenced, thus creating unique digital map images that allow the old maps to appear in their correct places on the modern globe. Some of the maps fit perfectly in their modern spaces, while others (generally earlier period maps) reveal interesting geographical misconceptions of their time and therefore have to be more distorted to fit properly in Google Maps and Earth. Cultural features on the maps can be compared to the modern satellite views using the slider bars to adjust transparency. The original historical maps are first made into digital images by scanning them with high resolution digital cameras. Then these digital images are transformed in a process called georeferencing, which makes them display in their correct geographical spaces in Google Maps and Earth. Georeferencing is done using a GIS program, which takes points on the old maps (cities, coast lines, rivers, streets) and connects them to the same points on a modern satellite map image or a modern street map or a modern map showing boundaries of countries and states. The GIS program then takes all these points (as many as 200 are made for very large maps) and uses them to recreate the digital image so it will fit into its modern geographical space. Often the image has to be curved a bit for this to be accomplished. The result is an exploration of time as well as space, a marriage of historic cartographic masterpieces with innovative contemporary software tools. The David Rumsey Map Collection was formed by David Rumsey over the last 30 years. Rumsey has scanned more than 17,000 of the maps since 1999 and put them online in a free public map library at www.davidrumsey.com. Rumsey is committed to free public access to his maps, so visitors can not only look at the maps in full high resolution (some of the digital map images are 2 GB) but they can also download them for personal use.

Electronic Cultural Atlas Initiative (ECAI)
URL: www.ecai.org
Location: International and Area Studies, UC Berkeley.
Sponsor(s):
Principal(s): Lewis Lancaster, Michael Buckland.

Period: Not explicitly historical, though many cultural properties are relatively static and therefore represent “historical” features of the landscape.

Area: Global.

Availability: Posted on-line.

Status: Ongoing.

Description: Map and data repository; portal to many cultural atlases, linking independent projects in which a GIS approach is taken to cultural properties of regions. ECAI also sponsors conferences, establishes standards for academic publishing with GIS data, and undertakes other scholarly activities.

Gallica
URL: http://gallica.bnf.fr/
Location: Bibliothèque Nationale de France.
Sponsor(s): Bibliothèque Nationale de France.
Principal(s):
Period: non-specific
Area: French exploration/colonialism.
Availability: Free, on-line.
Status: Ongoing.
Description: An extensive set of on-line maps (not digitized).

Gateway for Historical Geographical-Information Systems (his-GIS)
URL: www.his-gis.net
Location:
Sponsor(s):
Principal(s):
Period:
Area:
Availability:
Status: Begun in 2007. There is virtually nothing on the site at present.
Description: [From the web site...] At his-GIS.net you can publish after registration dates of events like exhibitions or conferences. The forum is open for academic discussion. Furthermore Weblinks can be listed and users can strengthen their academic network.

Global Land Precipitation Dataset
URL: www.cru.uea.ac.uk/~mikeh/datasets/global/
Locations: Climate Research Unit, University of East Anglia
Sponsor:
Principals: Phil Jones, Mike Salmon.
Area: Global.
Availability: Free. Data are posted.
Status:
Description: An historical monthly precipitation data set for global land areas from 1900 to 1998, gridded at two different resolutions (2.5° latitude by 3.75° longitude and 5° latitude/longitude) has been constructed and is available for use in scientific research.
Global Temperature Dataset
URL:  www.cru.uea.ac.uk/cru/data/temperature/
Locations:  Climate Research Unit, University of East Anglia
Sponsor:  
Principals:  Phil Jones, Mike Salmon.
Area:  Global.
Availability:  Free. Data are posted.
Status:  
Description:  Land air temperature anomalies on a 5° by 5° grid-box basis from 1900-2000.

Hyperhistory
URL:  www.hyperhistory.com/online_n2/History_n2/a.html
Location:  
Sponsor(s):  
Principal(s):  
Period:  
Area:  
Availability:  
Status:  
Description:  [From the web site...] HyperHistory is an expanding scientific project presenting 3,000 years of world history with an interactive combination of synchronoptic lifelines, timelines, and maps. A display from the simple to the more complex. The graphics in the Center Panel provide a first quick overview of general information linked to short texts in the right Text Panel. More comprehensive information is provided at the bottom of the text files with www links. (From the Shakespeare text file, for example, you can find a link to the complete works of Shakespeare, etc.) In the History section is a link to a book text with 250 chapters of world history, and in the science section the student can expand to learn about the History of the Theory of Matter, or the Relativity Theory, etc. Navigation is made easy because the permanent Menu Panel on the left allows anybody to switch effortlessly from one category to another. The separate Text Panel on the right allows instant display of scripts while retaining the context provided by the synchronoptic graphics in the center Main Panel. A Color Code distinguishing between science, culture, religion & politics helps to absorb with ease an overwhelming amount of information. The Menu Panel contains four domain buttons and one Option: Options includes sections which do not fit into the other domains. A section called Connections contains stories or graphics as interesting 'footnotes' to world history. “People” displays lifelines for over 1000 persons in science, culture, religion, and politics. (In addition there are special expanded sections for scientists, artists, writers, composers, famous women, and politicians accessible from the 'Options' panel). “History” displays timelines for the major civilizations with many text links. “Events” depicts events on a year to year basis. At the present time it reaches back from 2008 to 1770. Maps displays seven Periods of World History with many links to dozens of regional maps. Index leads to an alphabetical index for over 1000 persons depicted within the lifelines of the People's section. HHO is complemented with a comprehensive world history text, accessible from the Book Text button on the lower right. Over 3,000 files are interconnected throughout the site. In addition to that HyperHistory provides several hundred links to the world wide web. The growing site itself contains presently over 100 MB of images and text files, but individual gif files are kept small enough to allow for a quick display. HyperHistory Online (HHO) - which is visited by 3,500-7,000 people a day - functions as a companion to the seminal World History Chart by Andreas
Nothiger. So far over 35,000 charts have been sold worldwide. Online does not replace the printed chart but complements it: the printed Chart confers a superior overview of history whereas the Online version provides depth and connections to other web sites. The two versions combined form an integrated part of the World History Project - thus serving as a model of how the printed and electronic media can complement each other. A Web Event HHO is a landmark in web development, because people from around the world are encouraged to participate. Participation can take many forms, from supplying information that could be included to suggestions of good Hyperlinks. Imagine a symphony of links with as many pathways as there are participating minds. The most exciting contribution will be if people get inspired to build their own independent web pages about subjects we didn't cover. High standards have to be maintained of course and we will have to make decisions about which websites would be appropriate to link up to.

**International Networks Archive**

*URL:* www.princeton.edu/~ina/index.html  
*Location:* Sociology Department, Princeton University  
*Sponsor(s):* No current funding.  
*Principal:* Miguel Centeno.  
*Period:* Mostly contemporary.  
*Area:* Global.  
*Availability:* Freely available, when completed.  
*Status:* At the planning stage.  
*Description:* [From the web site...]
The purpose of the Archive is to assemble data sets relevant to empirical research on mapping the global web in a central location and to standardize them so the various indicators can be combined. Given the immense amount of work that defining a global web involves we argue for disseminating the raw data as widely as possible so as to recruit the largest possible number of collaborators. Specific project components include: Collecting various network data sets (e.g. communication, trade, tourism, policy issues, migration); Establishing a uniform format for these so that they can be combined in models; Making data publicly available on our website. We are specifically interested in global communication and information technologies, international inequality, and issues of international security. Project #1: Global Communication and Information Technologies. Project #2: International Inequality. This project would provide a map for understanding the composition and consequences of the global society by emphasizing the relational position of countries in a world of flows and exchanges.

**Mapping History Project**

*URL:* http://mappinghistory.uoregon.edu/english/index.html  
*Location:* University of Oregon, University of Munster  
*Sponsors:* Oregon University System, the University of Oregon, the Norman Brown Family Fellowship Fund, the Alexander von Humboldt Foundation, the Universitat Munster  
*Principals:* James Mohr (Oregon), John Nicols (Oregon), Peter Funke (Münster), Barbara Stollberg-Rilinger (Münster)  
*Period:*  
*Area:*  
*Availability:*  
*Status:*  
*Description:* A map repository; no data extracted.

**Open Street Map**
URL: http://en.wikipedia.org/wiki/OpenStreetMap
Principals:
Period: Contemporary.
Area: Global.
Availability: Free.
Status: Ongoing.
Description: A collaborative project to create a free editable map of the world. The maps are created using data from portable GPS devices, aerial photography and other free sources. Both rendered images and the vector dataset are available for download under a Creative Commons Attribution-ShareAlike 2.0 licence. Registered users can upload GPS track logs and edit the vector data using the given editing tools. Inspired by sites such as Wikipedia — the map display features a prominent 'Edit' tab and a full revision history is maintained.

The Plough
URL: 
Locations: Harvard University and UCLA
Sponsor: 
Principals: Alberto Alesina, Paola Giuliano, Nathan Nunn.
Period: 1000-present.
Area: Global.
Availability: Free. Data will be posted on-line once paper is published.
Description: Includes data on the historic use of animal plow agriculture. Uses the Ethnographic Atlas, and geocodes the location of the 1200 ethnic groups in the dataset.

Project on Religion and Economic Change (PREC)
URL: www.prec-online.com
Location: UT Austin.
Sponsors: Spiritual Capital project of the Metanexus Institute; the Templeton Foundation.
Principals: Robert Woodberry.
Area: Non-western.
Availability: Free, once completed.
Status: Ongoing.
Description: [From the web site…] The goal of the Project on Religion and Economic Change is to evaluate the effect of religion on economies around the world at the national, community, and individual levels. The central question is one that dates back to sociologist Max Weber and his study of Protestant beliefs and the rise of capitalism. Countless scholars have wrestled with the question, but have largely been frustrated by methodological or data challenges. One of the major obstacles has been a difficulty in finding credible, detailed historical information for cross-national comparisons. PREC is unique in its scope and design. Our data sets, based on missionary documents, permit an unprecedented look at the cumulative effect of religious influence on economies over 190 years and in every non-Western society. Innovative use of Brazilian census data and surveys in Malawi allow scholars a more detailed picture of possible religious effects. The findings produced by PREC will illuminate this age-old question with the best of modern social science, technology, and a sense of global perspective. Perhaps more importantly, the vast amount of raw data collected through this project provides researchers with an unparalleled new resource
help answer questions regarding the health, educational, legal, religious, and economic conditions of many nations across nearly two centuries.

**Timemap**  
**URL:** www.timemap.net  
**Location:** Archeological Computing Laboratory, University of Sydney  
**Sponsor(s):**  
**Principal(s):** Ian Johnson  
**Period:**  
**Availability:** Free  
**Status:**  
**Description:** A piece of software, not a historical project per se. [From the web site...] TimeMap TMJava is a novel mapping applet which generates complete interactive maps with a few simple lines of html. It provides a way of easily enriching web pages with historical or contemporary information that goes far beyond static jpg map images. It’s easy for beginners, yet provides completely customisable power and distributed backend database connectivity for the expert. It's free for personal use. TimeMap's unique time-handling provides an engaging and intuitive method of delivering historical, community, government, research and business information. Combining mapping and the time dimension gives new ways of visualising urban growth, the spread of empires, heritage sites, environmental change, weather patterns, traffic flow, earthquakes, mobile network faults, and much more — ranging in time scale from millions of years to seconds. TimeMap time-filters and animates maps on the fly, connects to datasets anywhere on the web and can search for and load thousands of local maps dynamically as you zoom and pan. TimeMap can filter huge datasets server-side and download only the data needed, or work standalone off a CD. It adapts legends dynamically as scale changes and generates hyperlinks on-the-fly between objects on the map and web pages, and is completely customisable with XML. Yet the applet weighs in at only 350K! TimeMap's Windows tool, TMWin, allows you to build and publish interactive maps on your web site without any programming skills. You can use your own data or map data from the Electronic Cultural Atlas Initiative's (www.ecai.org) data clearinghouse. For advanced users TimeMap can be extensively customised with XML and JavaScript. TMJava source code is available to members of the TimeMap Open Source Community (see under Community). The TimeMap project team can provide rapid response support to user, from neophyte to IT professional, as well as a full range of custom database and TimeMap programming services. The project is currently developing improved Flash animation output and a combined map/timeline display of historical events. Historical events are stored in our collaborative social bookmarking/bibliographic database, Heurist. Visit HeuristScholar.org and log in as guest + guest for a preview. This development will allow students, teachers and members of the public to build a cutting-edge historical event map/timeline simply by entering dates, descriptions and map locations in the database, and attach wiki-based commentary and threaded discussions to events. We are expanding our staff by two new positions to work on end-user support and application development, documentation and the much-delayed release of the Open Source version of TimeMap.

**World Language Mapping System**  
**URL:** www.gmi.org/wlms/  
**Location:** Colorado Springs, CO  
**Sponsor(s):** Global Mapping International (a Christian organization)
Principal(s):  
Period: Contemporary.  
Area: Global.  
Availability: Standard WLMS licensing permits use of the data at a single (or optionally, at multiple) GIS end-user workstations, and distribution of maps created on licensed workstations in either print or electronic form with only minor attribution requirements. See the standard license (in PDF format) for details. Custom licensing for large GIS sites, multi-user map servers, or other situations not covered by the standard license can be negotiated on a case-by-case basis. Cost for first work-station: $1,000 ($350 for educational institutions). For ordering questions, contact Bill Dickson, +1-719-531-3599 x592 or bill@gmi.org.  
Status: Ongoing.  
Description: [From the web site…] The World Language Mapping System consists of Geographic Information System data mapping language locations both as points and polygon (of language homelands), with attribute information from Ethnologue: Languages of the World, Fifteenth Edition. The World Language Mapping System (WLMS) is the result of over 17 years of collaborative work between Global Mapping International (GMI) and the SIL International (SIL), to map the over 6,800 languages described in SIL's 15th edition Ethnologue. The resulting GIS data is used by SIL to produce maps for the Ethnologue and other purposes, and has been published by GMI in GIS format for the Christian nonprofit community since 1995 as part of the Global Ministry Mapping System. This data set, adapted to work with either NGA's public domain Digital Chart of the World (VMAP Level 0) base map or GMI's companion Seamless Digital Chart of the World, is now available to the broader community of GIS users.

AFRICA

AfricaMap  
URL: http://cga-3.hmdc.harvard.edu/africamap/  
Location: Center for Geographic Analysis (CGA), Harvard University  
Principals: Ben Lewis  
Status: Ongoing.  
Period: Mostly contemporary.  
Availability: Most if not all of the data will be made public.  
Description: [From AfricaMap Release I, Beta. A Framework to Support Collaborative Research and Teaching, February, 2009. http://africamap.harvard.edu] Four broad categories of mapping will be included in AfricaMap: 1) Contemporary maps available in the Harvard Map Collection and elsewhere; 2) Human Area Relation File Data related to Africa (information on African cultures - population density, languages, subsistence, beliefs, technologies, housing forms); 3) Historical maps of the continent (these will be geo-referenced); 4) Scholarly maps focused on Africa in various historical periods. The project will form a framework for referencing an array of materials important for different areas and periods (photographs, charts, videos, music segments, spoken language exemplars, etc.).  

URL: www.brown.edu/Research/AAAH/index.htm
Location: Brown University
Principals: Nancy Jacobs, Rolando Peñate
Sponsor(s):
Period: 1879-2002
Area:
Availability:
Status: Probably not ongoing
Description: Data includes: Territory names; Boundaries, Imperial rulers; Political systems; Violent conflicts; Economic and demographic trends

The Long-Term Effects of Africa’s Slave Trades
URL: www.economics.harvard.edu/faculty/nunn/data_nunn
Locations: Harvard University
Sponsor:
Principals: Nathan Nunn
Period: 1400-present.
Area: Africa.
Availability: Free. Data are posted on-line.
Status: Published in the Quarterly Journal of Economics.
Description: Country level data on historic slave exports during the Indian Ocean, trans-Atlantic, Red Sea, and trans-Saharan slave trades each century from 1400 to 1900.

The Slave Trade and the Origins of Mistrust in Africa
URL: www.economics.harvard.edu/faculty/nunn/files/Trust_v6.pdf
Locations: Harvard University
Sponsor:
Principals: Nathan Nunn, Leonard Wantchekon.
Period: 1400-present.
Area: Africa.
Availability: Free. Data will be posted on-line once paper is published.
Description: Ethnicity level data on historic slave exports during the Indian Ocean and trans-Atlantic slave trades. Also includes information on the 19th century boundaries of ethnic groups within the African continent.

EAST ASIA

China Historical GIS Project (CHGIS)
URL: www.fas.harvard.edu/~chgis/
Location: Center for Geographic Analysis (CGA), Harvard University
Sponsors: Luce Foundation; Yenching Institute (Harvard University); Asia Center (Harvard University); National Endowment for the Humanities.
Principals: Peter Bol; Lex Berman, Man Zhimin.
Period: 222 BCE-1911.
Area: China.
Status: Ongoing.

Availability: To be determined. Most of the data will probably be made public.

Description: [From Bol (2007)…] CHGIS is a spatial database including the administrative hierarchy from the regime capital to the county (xian) level. At present, a varying number of towns (cunzhen) within the county are included for two points in time: 1820 and 1911. Counties and towns are given as point files. Prefectures and higher order units are given as polygons for the territory administered and point files for the administrative capitals. County polygons are also available for 1911. The core datasets are ‘time-series’ which include all administrative changes dated to the year. The 1820 and 1911 datasets are ‘slices-in-time’ and pertain to the administrative structure and major towns around those years. The project currently aims to cover the eighteen core provinces, home to 90% of the population, with the exception of the dynastic coverage of the Qing empire in 1820.

SOUTH ASIA

Digital South Asia Library (DASL)

URL: http://dsal.uchicago.edu/about.html
Maps: http://dsal.uchicago.edu/maps/
Location:
Sponsor(s): Association of Research Libraries' Global Resources Program, the Andrew W. Mellon Foundation, the Technological Innovation and Cooperation for Foreign Information Access (TICFIA) grant from the US Department of Education, the Ford Foundation (New Delhi), The Asia Society.
Participants: [From the web site…] Leading U.S. universities, the Center for Research Libraries, the South Asia Microform Project, the Committee on South Asian Libraries and Documentation, the Association for Asian Studies, the Library of Congress, the Asia Society, the British Library, the University of Oxford, the University of Cambridge, MOZHI in India, the Sundarayya Vignana Kendram in India, Madan Puraskar Pustakalaya in Nepal, and other institutions in South Asia. For a complete list see http://dsal.uchicago.edu/institutions.html
Principal(s):
Period:
Area: South Asia.
Status: Appears to be ongoing.
Availability: Material available for private use only; map images are downloadable but do not appear to be of high quality in this format.
Description: [From the web site…] “The Digital South Asia Library provides digital materials for reference and research on South Asia to scholars, public officials, business leaders, and other users.” Most of the materials appear to be non-cartographic; however, they have maps drawn from a few historical atlases, though these are not digitized and not suitable for digitization in current form (as posted on the web).

Railroads and the Raj: The Economic Impact of Transportation Infrastructure

URL: http://personal.lse.ac.uk/DONALD1S/donaldson_JMP.pdf
Locations: London School of Economics
Sponsor:
Principals: David Donaldson.
Period: 1800-present.
Area: India.
Availability: Free. Data will be posted on-line once paper is published.
Description: Includes Indian district level, price, production, and bilateral trade data at the commodity level, as well as daily climate data from nearly 3,600 rainfall stations within the continent from 1891 to 1930. Has digitized and geocoded India's railroad network every decade between 1860 and 1930.

EUROPE/ASIA MINOR/MIDDLE EAST

Belgian Historical GIS
URL:
Location:
Sponsor(s):
Principal(s):
Period:
Area:
Status:
Availability:
Description:

Centennia Historical Atlas (web-based and cd-rom)
URL: www.historicalatlas.com/centennia.html
Location:
Principals: Frank Reed
Period: 1000-present.
Area: Europe and Middle East. To be extended to cover the history of North America from the pre-Columbian period.
Availability: Would have to be negotiated.
Status: Improvements and additions ongoing.
Description: CD-rom. Primarily borders of countries/empires, we believe.
[From the web site...] A map-based guide to the history of Europe and the Middle East from the beginning of the 11th century to the present. It is a dynamic, animated historical atlas including over 9,000 border changes. The map controls evolve the map forward or backward in time bringing the static map to life. Our maps display every major war and territorial conflict displaying the status of each region at intervals of a tenth of a year. The maps reflect actual "power on the ground" rather than internationally-sanctioned or "recognized" borders.

Digital Atlas of Ancient Civilizations in Europe and Middle East
URL: medievalmap.harvard.edu
Location: Center for Geographic Analysis (CGA), Harvard University
Sponsor(s):
Principal(s): Michael McCormick; Guoping Huang.
Period:
Area:
Availability: Public, once published.
Status: Ongoing.
Description: Digitized data on roads, settlements, shipwrecks, administrative boundaries, etc.

**Euroatlas**
URL: http://www.euratlas.com/summary.htm
Location: Rue du milieu, 30, Yverdon, Switzerland
Sponsor(s):
Principal(s): Christos Nussli
Period: 1-2000 AD.
Area: Europe.
Availability: Free.
Status: Completed.
Description: [From the web site…] The Periodical Historical Atlas of Europe shows the evolution of this continent through a sequence of 21 historical maps, every map depicting the political situation at the end of each century. In the high resolution pages, you will find 21 minimaps giving access to 84 quarters of maps with detailed views of the states, provinces and cities. Moreover, for each century, a direct access to the complete map of Europe is provided.

**German Historical Geographical Information System (GHGIS)**

**Great Britain Historical Geographical Information System (GBHGIS)**
URL: www.port.ac.uk/research/ghgis/abouthistoricalgis/
Vision of Britain Through Time: www.visionofbritain.org.uk/index.jsp
Wikipedia entry:
http://en.wikipedia.org/wiki/Great_Britain_Historical_GIS#Is_it_a_GIS_at_all.3F
Location: University of Portsmouth
Sponsor(s):
Principal(s):
Period:
Area:
Availability:
Status:
Description: [From the web site…] 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, travellers' tales and historic maps, assembled into a whole that is much more than the sum of its parts. This site tells you more about the project itself and about historical GIS. A separate site, funded by the UK National Lottery, has been created to make this resource available on-line to everyone, presenting our information
graphically and cartographically. This site is called A Vision of Britain Through Time and presents
the history of Great Britain through places. Our original work was entirely focused on historical
statistics - on census reports, data on births, marriages and deaths, and on unemployment and
poor law statistics. The resulting system, developed in London between 1994 and 1999, loosely
linked a record of changing administrative boundaries, managed using ArcInfo software, to our
existing database which held millions of statistical data values in hundreds of tables. Since 2001,
with funding from the UK National Lottery, we have built a completely new system designed both
to remove limitations of the original system and to hold a much broader range of the content: Like
any mainstream GIS, the original GBH GIS could hold information only about units whose
locations we knew. There are a few historical units which appear, for example, in tax lists but
whose location is unknown. There are a great many more whose boundaries have yet to be
mapped. The core of our new system is a systematic list of all the units we know about - currently
over 48,000 units, linked by over 150,000 relationships. This core system is not, strictly speaking, a
GIS at all: it is implemented using Oracle database software, requires no locational data at all and
is organised as an ontology, or "polyhierarchic thesaurus". Each unit can have any number of
names, hierarchic relationships are held very flexibly, and we use a system of "date objects" which
enable us to record changes as precise calendar dates, as years, or as strings of text such as "at least
1174 but possibly as early as 983". Although knowing boundaries is not compulsory, we use the
Oracle Spatial extension to hold over 40,000 boundary polygons, with dates, for many units. These
polygons were created by our own earlier work, by Roger Kain and Richard Oliver's work at
Exeter University on the boundaries of Ancient Parishes, and recent work we have done on
Scottish parish boundaries. The system can use hierarchical relationships to infer approximate
locations for units lacking boundaries. We have built another quite separate GIS, using open
source MapServer software which holds three complete sets of geo-referenced scanned images of
Ordnance Survey one inch-to-the-mile maps of Britain, partly to put our boundary mapping in
context: the 19th century First Series, the 1940s New Popular Edition and the inter-war Land Use
Survey of Great Britain (this last set were funded by the Environment Agency and the Department
for Environment, Food and Rural Affairs). The system as a whole aims to be a comprehensive
description of Britain and its localities. One way we achieve this is by including approaching 10m.
words of text: of accounts of journeys around Britain, such as Celia Fiennes' Through England on
a Side Saddle in the Time of William and Mary, cross-referenced by place to the rest of our
system; over 90,000 entries from descriptive gazetteers published in the late 19th century,
describing towns, villages and landmarks; and the main report from every census up to 1961. A
large part of our statistical holdings are being moved to this new structure. Most collections of
computerised statistics, including our own old database, are divided into many separate tables and
are designed mainly to allow researchers to download these datasets for further analysis. Our new
system holds all statistical data values in a single column of a single table, with many millions of
rows, enabling them to be used very flexibly for mapping, graphing and reconstructing the source
table. That last feature is made possible by a metadata framework developed by the Data
Documentation Initiative. The extent of our support for open standards is probably unique among
historical GIS projects. We also support Dublin Core, the Alexandria Digital Library Gazetteer
Content Standard, and the Ethnologue and Linguist codes for identifying the languages of place-
names. Our creation of maps for the web follows Open GIS Consortium standards. The central
focus of the project is on the development of our GIS as a national resource. However, there have
been a series of associated major research projects for example, an analysis of the impact of high
unemployment in inter-war Britain on both short-run infant mortality rates and long-run life
expectancy. We also undertake consultancy work.
The post-2000 GB Historical GIS makes no use of commercial GIS software, except for editing parts of the content, and implements a data model which could not be implemented using packages such as ArcGIS or MapInfo, so is it a GIS at all? It is certainly not a conventional GIS, but one answer is that any system that can create an image like the one shown below is some kind of GIS. This image from *A Vision of Britain through Time* combines the boundaries of local government districts, data on unemployment from the 1931 census, and a scanned image of an Ordnance Survey ten mile-to-one inch map from the early 20th century. (GBHG GIS will be integrated into QVIS in the future.)

**Institute of European History Map Server**

**URL:**
**Location:**
**Sponsor(s):**
**Principal(s):**
**Period:**
**Area:** mainly Germany
**Availability:**
**Status:**
**Description:**

**Mapping Europe's Historic Boundaries and Borders**

**URL:**
**Location:**
**Sponsor(s):**
**Principal(s):** Michael Goerke; Humphrey Southall
**Period:**
**Area:**
**Availability:**
**Status:**
**Description:** Conference held in Florence/Firenze, Italy in June 2000.

**Maps of War/Imperial History of the Middle East**

**URL:** www.mapsofwar.com/images/EMPIRE17.swf
**Location:**
**Sponsor(s):**
**Principal(s):**
**Period:** 3000BC-present
**Area:** Middle East
**Availability:**
**Status:**
**Description:** Nifty Flash presentation showing changing boundaries of empires through time.

**Portuguese National GIS**

**URL:**
**Location:**
**Sponsor(s):**
**Principal(s):**
**Period:**
Query and Context based Visualization of Time-spatial Cultural Dynamics (QVIZ)

**URL**: www.qviz.eu/index.php

**Locations**: Various.

**Sponsors**: European Union.

**Principals**: Patrick Svensson, Fredrik Palm (both at Umeå University). [From the web site…] QVIZ has seven partners, from the countries Austria, England, Estonia, Spain and Sweden. The network is coordinated by Umeå University in Sweden. The National Archives of Estonia and Sweden provide the archival expertise in the project as well as the highly structured content needed to develop the system. Salzburg Research Forschungsgesellschaft contributes the knowledge and technology necessary to build the collaborative environment that enables the social knowledge building functions of the system. The Map and Faceted Query browser is essentially an amalgamation of the knowledge from three partners. The map component was developed by the Estonian company REGIO and adapted for the use in the QVIZ-portal. The faceted browser has been developed by Umeå University. Both the map component and the faceted browser are dependent on the administrative unit ontology provided by the University of Portsmouth that enables the system to follow the changes of the respective unit borders over time. The Spanish company Telefónica provides the project with high-level design competence and technical specifications for the system.

**Period**:

**Area**: Europe.

**Availability**: Presumably free, once completed.

**Status**: Begun in 2006; ongoing.

**Description**: [From the web site…] QVIZ is a project started in order to bring users a single entry point to the archives of Europe. A common starting point that allows browsing of the archival resources through time and space using a dynamic map or contextual categories. The map interface locates the resources without the need of knowledge concerning the language that the resource keeps or which institution that holds the records. In addition to this, QVIZ also provides an environment for collaborative knowledge building with social bookmarking for interested users. Many archives are working on digitalizing material in order to enhance access and preserve the fragile content in the archives. Most existing digital archives do not allow the user to add or bookmark information in the system. However, the information that the archives hold is often complex, which impedes the possibilities for an effective and facilitated search. As the amount of digital material steadily increases, an easier way to conduct investigations of the archives becomes a necessity. During the first year the activities focused on underlying research, constructing a technical framework, system requirements and specifications as well as prototype design. In the second year, the focus changed to integration of the different parts of the platform. QVIZ is currently working on its third integrated prototype which involves fine-tuning the knowledge models and integral environments.

**NORTH AMERICA**
National Historical Geographic Information System (NGHIS)

URL: www.nhgis.org

Location:
Sponsor(s): NSF.
Principal(s):
Period: 1790-2000
Area: US.
Availability: Free, on-line.
Status: Ongoing.

Description: [From the web site...] Aggregate census data and GIS-compatible boundary files for the United States. Options: (1) The NHGIS Shape Finder provides access to boundary files for mapping data in a GIS software package. (2) The NHGIS Data Finder provides access to aggregate census data. (3) Social Explorer allows online mapping of NHGIS data.

Quantifying the Long-Term Impact of Environmental Change: Evidence from the American Dust Bowl

URL: http://econ-www.mit.edu/files/3551

Locations: MIT
Sponsor:
Principals: Richard Hornbeck.
Area: USA.
Availability: Free. Data will be posted on-line once paper is published.

Description: Includes digitized and geocoded data on soil erosion during the dust bowl of the 1930, as well as economic, agricultural, and social outcomes at the county level.

LATIN AMERICA

The Persistent Effect of Peru’s Mining Mita

URL: http://econ-www.mit.edu/files/3144

Locations: MIT
Sponsor:
Principals: Melissa Dell.
Period: 1573-1812.
Area: Peru and Bolivia.
Availability: Free. Data will be posted on-line once paper is published.

Description: Includes digitized and geocoded data on the location of the mita, and of subsequent social and economic outcomes at the district level.
Appendix C:
Historical Atlases

Canada

United States

North America

Central America

South America

Latin America

Central/Eastern Europe
Europe

Balkans

Middle East

Russia/Central Asia/Caucasus

Africa

**South Asia**

**Southeast Asia**

**Asia**

**The Pacific**

**British Empire**

**The World**
*Historical Atlas of the 19th Century World 1783-1914*.


**Exploration**


**Military**


**Railroads**


**Cultures**


**Religion**


*Historical Atlas of Religions*, a four-volume set published by Facts on File.


**Christianity**

**Islam, Muslims**

**Transport**

**Language**
Appendix D:  
Non-Cartographic Data Sources

General
Duignan, Peter; L.H. Gann. Colonialism in Africa (5 Volumes). Hoover Institution Publications.


Europa Yearbook, 1926-.


Gapminder. In process. Historical data on infant mortality and life expectancy (mostly derived from James Riley’s original dataset). Stockholm, Sweden.

Gerring, John; James Mahoney. 2007-11. “Colonialism and Its Legacies: A Comprehensive Historical Dataset.” [NSF-sponsored project for the collection of data pertaining to colonialism and long-term development.]

Global Financial Data (GFD). A proprietor. (www.globalfinancialdata.com/) Includes: Equity data back to 1693, Fixed income data back to 1694, Exchange rates back to 1590, Inflation data back to 1271, Commodity data back to 1257, Total returns on stocks, bonds and bills back to 1800, U.S. stock data back to 1815, Real estate data back to 1830.


Housenick, Christopher. In process. Dataset on transportation systems (railroads, automobiles, and air travel) and communication systems (telegraphs, telephones, cellular services, and the Internet), 1840-2006.

International population census publications, series II, pre-1945: Africa.


Maddison, Angus. 2008. [Data for population, GDP, and GDPpc from 1 AD to the present.] http://www.ggdc.net/Maddison/content.shtml

Manning, Patrick. In process. Demographic historical data for Africa.


McGowan, Patrick. *British Economic Imperialism, 1869-1914.* ICPSR--Study No. 7738
Mitchener, Kris James; Marc Weidenmier. 2007. “Trade and Empire.” Unpublished manuscript. [Associated dataset.]
*Statesman's Yearbook*, 1864-.
The Making of the Modern World. Digital facsimile of 61,000 works of literature on economics and business from 1450 through 1850; combines the Kress Collection of Business and Economics at the Baker Library, Harvard Business School and the Goldsmiths' Library of Economic Literature at the University of London Library; coverage: 1450-1850.
*Union Postale Universelle Statistique Generale*, 1875-. Bureau International de L’Union postale universelle.
Wilkinson, Steven I. In process. *Colonization, Institutions and Conflict*. Book manuscript. [Associated dataset.]

**British Empire/UK**

*British and foreign state papers*. Covers 1812 to c.1967.
*Colonial Statistical Tables Command Papers*.
*Great Britain Colonial Office Reports (“Blue Books”).* 1886-.


*Statistical Abstract for the Several Colonial and Other Possessions of the United Kingdom, 1874-1901*.

**French Empire/France**

*Documents diplomatiques francais*. Ministere des affaires etrangeres, Commission de publication des documents francais.

French Colonial Office. 1888- *Annuaire Colonial*. [Each colony publishes its annuaire, a summary of the administrative organization with lists of officials.]


**German Empire/Germany**

*Deutscher Kolonial-Kalendar*

**Ottoman Empire/Turkey**


**Portuguese Empire/Portugal**

**Spanish Empire/Spain**


**American Empire/United States**


United States Dept. of the Treasury. 1901. *Colonial Administration, 1800-1900: Methods of Government and Development Adopted by the Principal Colonizing Nations in Their Control of Tropical and Other Colonies and Dependencies*. Published by Govt. Print. Off.

**Countries**

**Algeria**
- Service de la Statistique Générale: *Statistique Générale de l’Algérie* 1867-1925
- Service Central de Statistique: *Annuaire Statistique de l’Algérie* 1926-1964

**Angola**
- Repartiçao de Estatística Geral: *Annario Estatístico* 1933-1973

**Benin**
- Institut National de l’Analyse Economique: *Annuaire Statistique* 1965-1975

**Botswana**

**Burundi**
- Département des Etudes et Statistiques: *Annuaire Statistique* 1969-1975

**Cameroon**
- Direction de la Statistique et de la Comptabilité Nationale: *Note Annuelle de Statistique* 1973-1975

**Cape Verde Islands**
- Secção de Estatística Geral: *Annuario Estatístico* 1933-1952

**Central African Republic**
- Direction de la Statistique et de la Conjuncture: *Annuaire Statistique de la République Centrafricaine* 1962

**Chad**
- Sous Direction de la Statistique et des Études Economiques: *Annuaire Statistique* 1966-1975

**Congo (formerly French Congo)**

**Egypt (See also United Arab Republic)**
- Maslahat al-Ihsa wa-al-Ta’lad: *Statistical Returns* 1881-1897
- *Annuaire Statistique* 1901-1959

**Ethiopia**

**French Equatorial Africa**
- Haut Commissariat: *Annuaire Statistique de l’Afrique Equatoriale Française* 1936-1955

**French West Africa**
- Direction des Services de la Statistique Générale et de la Mécanographie: *Annuaire Statistique de
l'Afrique Occidentale Française 1949-1954

Gabon

Gambia
Statistical Summary 1964-1968

Ghana (formerly Gold Coast)

Guinea-Bissau (formerly Portuguese Guinea)
Repartição Provincial dos Serviços de Economia e Estatística Geral: Annario Estatístico 1947-1958

Ivory Coast
Ministère du Plan: La Côte d'Ivoire en Chiffres: Annuaire Statistique de la Côte d'Ivoire 1975

Kenya (formerly British East Africa)

Lesotho (formerly Basutoland)

Libya
Census and Statistical Department: Statistical Abstract 1958-1974

Malagasy Republic (Madagascar)
Service de Statistique Générale: Annuaire Statistique de Madagascar 1938-1951

Malawi (formerly Nyasaland)
Malawi Statistical Yearbook 1972-1974

Mali (formerly French Sudan)
Service de la Statistique Générale de la Comptabilité Nationale et de la Mécanographic: Annuaire Statistique de la République du Mali 1963-1973

Mauritania

Mauritius
Quarterly Digest of Statistics 1961-1966
Bi-annual Digest of Statistics 1966-1976

Morocco
al'Maslahah al-Markaziyah lil-Ihsa 'iyat: Annuaire Statistique du Maroc 1925-1976

Mozambique
Repartição Técnica de Estatística: Annario Estatístico/ Annuaire Statistique/ Statistical Yearbook 1926-1973

Niger
Direction de la Statistique: Annuaire Statistique 1962; 1967

Nigeria
Eastern Region. Statistics Division: Annual Statistical Digest 1963-1965
Northern Region. Ministry of Economic Planning: Statistical Yearbook 1964-1966

Senegal
Sierra Leone

Somalia
Central Statistical Department: Koobaba Istatistikada/ Statistical Abstract 1964-1973

South Africa
Department of Statistics: Statistical Year Book 1964-1966
South African Statistics 1968-1976 (Sheehy CB210)

Spanish Sahara
Secretario General: Sahara Espanol: Anuario Estadistico 1948-1950

Swaziland

Tanzania (formerly Tanganyika and Zanzibar)
East Africa High Commission. East African Statistical Department. Tanganyika Office:
Statistical Abstract 1938-1970

Togo
Direction de la Statistique: Annuaire Statistique du Togo 1966-1973

Tunisia
Direction des Affaires Economiques: Statistique Générale de la Tunisie 1913-1939
Service des Statistiques: Annuaire Statistique de la Tunisie 1940-1971

Uganda

United Arab Republic

Zaire (formerly Belgian Congo)
Statistiques Relatives à l'Année . . . 1957 and 1959

Zambia
Central Statistical Office: Statistical Yearbook 1967-1971 (Sheehy CG244)

Zimbabwe (formerly Rhodesia)
Department of Statistics: Official Yearbook of the Colony of Southern Rhodesia containing General Information and Statistics no. 1, 1924-no. 4, 1952
Statistical Yearbook of Southern Rhodesia: The Official Annual of the Social and Economic Conditions of the Colony 1938 and 1947