Proposal:
A Global Historical GIS (GH-GIS) Project

Principal Investigators (co-PIs)

Peter Bol – Professor, East Asian Languages & Civilizations; Director, Center for Geographic Analysis, Harvard University <pkbol@fas.harvard.edu>

John Gerring – Professor, Department of Political Science, Boston University <jgerring@bu.edu>

Nathan Nunn – Assistant Professor, Department of Economics, Harvard University <nnunn@fas.harvard.edu>

Humphrey Southall – Reader, Geography, University of Portsmouth <Humphrey.Southall@port.ac.uk>

Robert D. Woodberry – Assistant Professor, Department of Sociology, University of Texas, Austin <bobwood@austin.utexas.edu>

Collaborators

Harvard:

Wendy Guan – Director, GIS Research Services, Center for Geographic Analysis

Ben Lewis – Senior GIS Specialist, Center for Geographic Analysis

Portsmouth:

Paula Aucott – Senior Research Associate, Geography

Peter Collier – Lecturer, Department of Geography

Alex von Luenen – Senior Research Associate, Geography

John Westwood – Research Associate, Department of Geography

UT Austin:

Juan Carlos Esparza Ochoa – Profesor Titular, Departamento de Estudios Socioculturales, Instituto Tecnologico y de Estudios Superiores de Occidente, Mexico
For decades, geographic information system (GIS) techniques have been applied to problems in environmental and land management, military logistics, and geography. Only recently have scholars begun to appreciate its potential for other fields, and in particular for charting the course of human events through secular-historical time.¹ 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. Such projects typically cover only a few selected topics and countries, situated mostly in the developed world (e.g., in Germany, Great Britain, Canada, and the United States).²

The time seems ripe for the development of a truly Global and Historical GIS project (hereafter, GH-GIS), one oriented to serve the needs of scholars across the social sciences and humanities. Before proceeding to a detailed discussion of how such a project might be implemented, let us focus on some of the potential payoffs it might bring.

Broadly stated, the aim of historical GIS is to unify the world of spatial representation with the world of substantive information such that the two are interwoven seamlessly through time. That is, information about a topic (e.g., passenger traffic on a railroad line) is joined with a coordinate-based reference showing the location of that topic (e.g., the location of the railroad line), and traced over the course of days, years, or centuries.

The project envisioned here will allow for the storage and retrieval of any species of information that can be identified with a place and an approximate time-period. This includes text, numbers, pictures, maps -- all of which can be assigned a geographic location for use in GIS. Note also that each of these kinds of information may be stored in a variety of formats. Because file formats and the encoding of information at present varies so much it can be difficult to access and particularly to integrate files from many sources. GH-GIS promises a common framework to store and/or retrieve this information, as described below.

As part of the initial phase of the project, we intend to collect a large quantity of readily accessible information of importance to those working in the humanities and social sciences. Our expectation is that users will add to this collection, in Wiki fashion, so that the contours of GH-GIS change over time in response to the interests of users and the availability of new materials. There is ultimately no limit to the temporal scope or substantive focus of information that might be integrated into GH-GIS.

Note that while a spatial approach to knowledge is second-nature in certain fields, e.g., archeology, geography, and forensics, it is grossly under-utilized elsewhere. This is partly a product of the predominantly non-spatial ways in which we are accustomed to presenting information, e.g., in tables, datasets, statistical models, or in prose. Analysis follows format, and since the format is non-spatial the mode of analysis featured in most humanities and social sciences fields is non-spatial. It is hoped that by making information readily available in a spatial format GH-GIS will stimulate greater awareness of geographic concepts and spatial relations, as well as an improved grasp of the nature of geographic information.

A related problem is that scholars in many fields lack the technical capacity to grapple with their chosen topics in a spatial format. Few scholars in the social sciences are trained in the use of ArcGIS, the most commonly used software for GIS applications. Moreover, the information of substantive interest – be it political, economic, social, or cultural – is unlikely to be available in GIS format, requiring a time-consuming process of data-entry and digitization before any maps can be

² China is the one country in the developing world to have generated a substantial historical GIS project (Appendix A).
generated. We propose to make available a vast storehouse of information in GIS format – free to all users -- and to facilitate the process of adding new material to this compendium of (spatially organized) knowledge. We suspect that once this capacity becomes available, GIS will become a standard technique for those studying anthropology, architecture and design, art, history, language, literature, music, religion, and other humanities fields -- not to mention the social sciences.

Consider a topic such as the rise of a new religious movement. For such a topic, one will presumably want to pinpoint the origin of the movement and its spread through time. One may also want to map the life-histories of the founders of the movement (which is easily accomplished if biographical information on these individuals is available). One may want to disaggregate the topic by looking at the diffusion of particular ideas regarded as components of the general movement; this can be accomplished by noting the appearance of these ideas at different places and times. One may want to compare the diffusion of the movement with other movements (religious, political, military, and so forth) occurring in the same historical era, or in different eras. One may be interested in the way in which the diffusion of a religious movement follows extant trade routes, observes climatic zones, or adheres to particular types of terrain. Does the movement precede military conflict, or follow military conflict? Does it spread more rapidly across areas dominated by monotheistic religions, or animist religions? In what ways does the movement incorporate elements of the preceding religion?

Historical projects such as these dovetail with a renewed interest in world history and geography. Note that anything that can be observed about the human world may also be mapped, and GH-GIS promises to enhance the ease, flexibility, and sophistication with which maps on varied subjects (organized through time) can be created. Often, the visualization of spatial relationships adds something fundamentally new to a well-worn subject. At other times, researchers may wish to delve into relationships that are not apparent to the naked eye, or they may wish to give a more precise rendering to complex spatial relationships. In this vein, it should be noted that the GIS technology opens the way for mathematical modeling of spatial relations including network analysis, spatial regression, and other forms of spatial analysis.

GH-GIS will also enable those unfamiliar with ArcMap or other GIS software to create a wide variety of user-generated maps that are of publication quality. These may encompass the entire globe or may focus tightly on a chosen area (e.g., a country or region). Maps may single out a particular type of information (e.g., population or infrastructure), or may combine several layers (with separate color coding). Maps may focus on a single moment in time, on an extended time-period (averaging the data across the chosen period), or on changes through time (in moving picture format). The key point is that users from any discipline will be able to create maps that show the spatial dimension of their research, a task which would otherwise be beyond the capability of those without a working knowledge of ArcMap.

Another expected pay-off to GH-GIS is the interconnections that it allows – indeed, encourages – across fields. The contemporary humanities and social sciences are often criticized for their compartmentalization. Scholars tend to work in highly specialized areas, communicating only with others within their restricted sphere of expertise. This is partly a matter of habit. But it is also a product of the limited technological tools with which we work and the general problem of accessing diverse types of information that may be relevant to a given subject. Note that this information may consist of text, pictures, sound, video, or data (numbers), and each of these formats may be stored in any number of ways. Data, for example, may be stored as tab-delimited files or within any number of software packages. Accessing these files is hard, even for the initiated, and combining them with information from other sources involves additional hurdles. GH-GIS promises a common framework within which knowledge from different sources can be fruitfully accessed. Whatever sort of information one is working with, it can be mapped. This provides a common denominator across
which scholars from diverse fields and methodologies will be able to access information, converse with one another, and collaborate on cross-disciplinary projects.

Likewise, GH-GIS will encourage a three-dimensional approach to research: through time, through space, and across subjects. One can “dig” through (a) layers of information relevant to a geographic region (facilitating comparisons across subjects normally reserved to separate disciplines), (b) backwards and forwards through time (facilitating a comparison across historical eras), and (c) across adjacent regions (facilitating cross-region or cross-continent comparisons). When the various sources of information are properly linked, the researcher can make these comparisons visually and/or statistically. This not only enriches our understanding of specific subjects but also allows knowledge to cumulate successfully across fields. New information about a topic is, quite literally, layered across extant information as it is added to GH-GIS, for future scholars to encounter.

For policymakers and those engaged on the practical end of politics, we also envision pay-offs. Note that 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 is justified. Our assumption, however, 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 reasonably judge the issue until such an accounting has been accomplished. Indeed, for most topics of global interest -- including poverty, disease, premature mortality, social conflict, autocracy, poor governance, environmental degradation, and globalization -- a strong prima facie case can be made for the importance of history. The project envisioned here should serve to enhance the relevance of history for challenges facing the world today by making history accessible to those who are not professional historians.

In sum, the extraordinary flexibility afforded by GIS technology suggests that this project will become a basic resource for teachers and scholars in years and decades to come. Insofar as Google Earth serves as a foundation for geographical explorations of contemporary data, GH-GIS may come to serve as a principal entrée for contemporary and historical data pertaining to the humanities and social sciences. (Naturally, if it reaches this level of traffic, technical capacity will have to be expanded accordingly.)

We now proceed to a more detailed discussion of the project. This concerns (I) the gazetteer, (II) data (beyond time and space), (III) Colonialism and its Legacies (a prior NSF grant upon which this project builds), (IV) dataset construction, (V) infrastructure and applications, (VI) implementation and sustainability, and (VII) references. Additional details are contained in several appendices, which review (A) the architecture for GH-GIS, (B) background on relevant projects previously undertaken by the research team, (C) related projects undertaken elsewhere, (D) historical atlases of potential use to the project, and (E) additional data sources of a non-cartographic nature.

I. Gazetteer

A global gazetteer, i.e. a systematic listing of named locations, is an essential part of the project. Three distinct components of the gazetteer may be usefully distinguished. Firstly, users need to be able to find the places they are interested in, so we will include a large place-name gazetteer based mainly on
the GeoNames gazetteer (www.geonames.org) which contains over 8 million entries and itself draws heavily on the National Geospatial Intelligence Agency gazetteer. Neither of these sources contains dates or source information, so this initial place-name gazetteer will be only a starting point for building a global historical gazetteer.

Secondly, we will build a closely linked gazetteer of administrative areas which will be initially smaller, focusing mainly on states and quasi-states, and representing locations by polygons not points. This component will certainly record the changing peacetime boundaries of all sovereign and semi-sovereign bodies (including colonies and protectorates), as well as first-level administrative units (for non-microstates) from 1815 to the present, providing a continuous record of which political units existed, what they were called and where their boundaries lay. However, our aim is to extend the project back before 1815, and to cover contested boundaries and quasi-states. The construction of this resource will be a significant piece of historical scholarship, requiring dates and sources for all items of information, and will be the main task of the team at Portsmouth (and therefore the main use of AHRC funds). By including the changing boundaries of states, we can immediately make the place-name gazetteer more historical; for example, GeoNames places "Breslau" simply in Poland, but our system will also record that until 1945 it was in Germany, and before that in Prussia.

Thirdly, we will create tools which enable users to make the place-name gazetteer more historical, and the administrative unit gazetteer much larger. This part of the project will build on the considerable experience of the Great Britain Historical GIS team – in continuous existence since 1994. The original GBH GIS was a relatively conventional ArcGIS-based system for mapping census data by county, district and parish. From 2001 to 2004 it was rebuilt using a new architecture placing far greater emphasis on recording multiple names for areas and allowing name-based searching. In 2006-9 the system was again rebuilt, partly with European Union funding under the QVIZ project ("Query Visualisation"; www.qviz.eu), to provide a multi-national and multi-lingual resource (Aucott et al 2008; Aucott et al 2009, in press). The software infrastructure required is almost identical to that used by Africamap, as described in section V (below), and will provide the foundation for the proposed global gazetteer. The project will therefore benefit from an already mature information architecture, and – as we will be using staff who worked on the GBH GIS and QVIZ systems – we will be able to make a running start on gazetteer construction.

Note that the QVIZ system was concerned initially with the administrative geographies of Britain, Estonia and Sweden down to the parish-level. However, it already includes an initial listing of the states of Europe since 1815 (including subsequent international boundary changes). This will be a useful starting point for the new global gazetteer, though the software and data models developed for the earlier project are ultimately more important. Most existing digital gazetteers emphasize quantity, meaning sheer number of entries, over quality; this is true even of the actual digital gazetteer created by the Alexandria Digital Library (ADL) project, despite the very rich Gazetteer Content Standard (GCS) they have created (Alexandria Digital Library Project, 2004; Hill, 2006). The QVIZ system was designed to make full use of the ADL GCS, including multiple names in multiple languages, polygonal boundaries as well as point coordinates and, crucially, extensive chronological information. In one respect our data model departs from the ADL standard, because it does not require any coordinate data at all for an individual entry, but does require at least one hierarchical relationship. This reflects experience with actual historical sources, where hierarchical lists of named entities often exist prior to

3 Particularly large problems arise in a multi-national historical GIS where national boundaries change and one system of local administration replaces (or, more usually, overlays) another, so very clearly identifying these changes is an essential foundation for future elaboration. Sub-national units will be included partly based on cultural and political significance; for example, Cornwall is unique among English counties because it has its own language, an independence movement of sorts and an involvement in multi-national pan-Celtic structures.
accurate maps of the same entities (e.g., tax lists which locate townships within counties). This flexible structure allows footprints to be inferred for units (without explicit coordinate data) from relationships with entities that do have such coordinates. This means that the ADL Gazetteer Service Protocol can be supported, though the architecture is also designed to support other data standards such as the Open Geospatial Consortium's Gazetteer Service Profile and the archive sector's Encoded Archival Context (EAC) standard.

Because we identify geographical entities, together with the different names attached to each through time, rather than simply listing large numbers of names and coordinates, and because our entities are linked by a number of different kinds of spatial relationships, our architecture defines a true ontology. As such, our approach is not unique (see, for example, Kauppinen et al, 2008), but while most such structures have been developed by computer scientists as an adjunct to software projects, the Portsmouth team are all primarily historical researchers, for whom this approach provides a very flexible vocabulary for expressing historical complexity.
Figure 1: Simplified Gazetteer Data Model

A full entity-relationship diagram of the QVIZ system appears in Aucott et al (2008). Figure 1 simplifies this to the main tables holding gazetteer data (rather than metadata). The units table defines all administrative entities but holds very little information (an identifier, a type, dates of creation and abolition, and the sources of the information), as information held here cannot change. Although we already have a typology that defines “states” and “nations,” a good deal of additional conceptual and empirical work will be needed in order to satisfy the requirements of a project covering all states and quasi-states for the last two centuries. The relations table holds hierarchical relationships of different kinds (subordination within an empire; membership of multi-national bodies) but also textual information recording boundary changes (notably from Hertslet, 1875). Relations again have start and end dates. The status table holds more detailed information about the legal characteristics of entities.4 The names table can accommodate any number of names for each unit, with sources, dates and status; modern and “dead” languages are identified by combining the Ethnologue (www.ethnologue.com) and Linguist (http://linguistlist.org) coding schemes.

The above four tables narrowly define the ontology of administrative units, but the remaining two give it much of its power. The footprints table holds boundary polygons: a unit can have any number of polygons (or, strictly, multi-polygons where the unit involves detached portions or islands), distinguished by dates. One research question we will address is how best to include contested boundaries, which will involve including additional information here. The inclusion of coordinate data directly within a relational structure means any implementation requires an object-relational database. This architecture has already been implemented in both Oracle and Postgres, and the combination of spatial and semantic relationships has proved capable of making excellent use of as much or as little data as are available. Lastly, the places table holds the broader placename gazetteer already outlined. Although there are more places than administrative units, units are defined as belonging to places. This is because a whole series of units can be based on a single place, such as “London” or “Estonia.” The places table includes either a point coordinate, or a polygon (for bigger places like Estonia). Because there is just one names table, this simplifies searching and means that the whole structure is integrated within a single gazetteer.

4 Unlike the type held in the units table, status can change over time, and one research question is whether we can drop the former completely. This poses problems in creating conventional mapping and in supporting the ADL standard, but may enable a more historically appropriate treatment of entities like Estonia: a sequence of intermittent states, plus provinces in other people’s empires, but also a nation with a clear continuing history.
One reason for adopting this architecture is so that we can directly incorporate information on the legal status and relationships of states and quasi-states, and an important aspect of the project will be extending our taxonomy of units to cover states and quasi-states with greater precision. The following classification of (nonexclusive) categories provides our point of departure (but will surely be refined as we proceed): (a) legal sovereignty (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/protektorat/territory (a territory without internationally recognized sovereignty that belongs to another political unit to which it is not fully integrated), and (d) undefined territory (territory whose sovereign status is unclear or actively contested, i.e., where no authority enjoys an effective monopoly of force). For (a), (b) and (c) it will be important to identify the dominant unit. The principal source for legal sovereignty (a) is the Correlates of War (COW) dataset (www.correlatesofwar.org/). The principal source for effective sovereignty (b) is Adams (in process). Sources for hierarchical relationships (colony et al.) include Henige (1970), Hensel (various years), Strang (1990, 1991), and the Project on Religion and Economic Change (PREC) (Pl: Robert D. Woodberry) - described in Appendix B - and Colonialism and its Legacies (co-PIs: John Gerring and James Mahoney), described below.

A key feature of this architecture is that we can hold information, and provide inferred approximate locations for units with unknown boundaries and locations. Uncertainty and imprecision is inherent in historical sources, and this is also true of dates. We already implement a system of date objects which can hold anything from a calendar date to just a year or even just a text string, e.g., “around the reign of Edward II.” Further work is needed on methods for working with these objects, and especially with multi-lingual strings. This will build on existing research into time period thesauri (Petras et al, 2006; Timelines Thesaurus). Here again we must emphasize that while constructing this system requires a high level of technical competence, the intellectual challenges are ultimately interpretive and theoretical in nature, e.g., working through various ways of defining sovereignty (Krasner, 1999).

Work on administrative units, especially states, has been emphasized as this will be the main focus of our substantive research. Converting GeoNames content into our initial placename gazetteer will be straightforward, the main change being the removal of GeoNames entries for administrative units. The challenge lies in developing procedures for historically enhancing our existing placename gazetteer (developed in the course of the QVIZ project) while retaining the capacity to import additional content from GeoNames. A minimum requirement is that our place identifiers retain consistency with GeoNames. However, our tools will enable users to add additional names with attributes beyond those held in GeoNames, i.e. dates, sources and extended language codes.

One source for adding historical names will be historical maps. Work for the QVIZ project included the creation of a seamless map of Europe at 1:500,000 scale, based on British military (General Staff Geographical Survey) mapping from the 1940s. This is the largest scale at which the whole of Europe has been mapped within the space of a few years, meaning that the maps are comparable to one another and hence suitable as a baseline. More recently, the Imperial War Museum donated a large set of GSGS maps at the same scale covering much of the rest of the world. The Portsmouth team also has a good working relationship with the provider of maps to the UK military, the Defence Geographic Centre, who have offered to help identify additional map coverage, suitable to our purposes. British military mapping does not cover the Americas, so we will also draw on US Army Map Service (AMS) mapping, produced during and after World War II, which will require some off-site scanning elsewhere in the UK and in the Library of Congress. By focusing on relatively recent mapping, we will extend our seamless digital map of Europe (http://www.visionofbritain.org.uk/maps) into a map of the world showing, for example, railways at
roughly their greatest extent and cities before most modern growth. Longer term, there is considerable scope for expansion. For example, a related project at Portsmouth is currently creating a database of all British military mapping 1870-1980, and we have access to the Admiralty Library in Portsmouth which holds systematic mapping of coastal areas by the Royal Navy from the late 1700s on.

While collecting maps of the world we plan to create an archive of historical boundary mapping, gathering together digital images gathered in the course of our research from maps at all sorts of scales and dates. The vector boundaries held within our “gazetteer” will contain our best interpretation of changing historical borders, but this separate library of scans will allow other scholars to review and challenge our interpretations, and to add other more detailed geographies.

II. Data

Building on the gazetteer, we plan to collect and present information about several substantive historical topics. These include (1) governments, (2) demographics, (3) infrastructure, (4) slavery, (5) religion, and (6) conflicts. Because information on these subjects is widely available, spatially specific (tied to specific locations), generally comparable (through time and across space), and relevant to many substantive and theoretical concerns, they form an ideal empirical foundation for our project.

1. Government information includes (a) number of administrators or personnel, (b) revenue, and (c) expenditure. Information will be collected for various political units (e.g., imperial, national, and perhaps subnational) as well as for various departments and subdivisions (e.g., military, non-military, police, treasury, postal, transport,…) at the national level, as available.


2. Demographic information includes (a) population, (b) population density, (c) raw mortality (per 1000), (d) infant mortality (per 1000 live births), (e) life expectancy, and (f) births (per 1000 inhabitants). Each of these statistics will be collected for various spatial units (e.g., empires, countries, colonies, cities,…) and for various subgroups (e.g., male/female, ethnic groups, religious groups, missionaries, racial groups, settlers, indigenes, citizens/noncitizens, slave/free), as available.

For life expectancy and infant mortality, key sources include the datasets compiled by James Riley (available from Gapminder) and Peter Lindert, which include data across many countries and colonies in the developed and the developing world, extending back to the nineteenth century. For crude birth and death rates, key sources include Murphy and Dyson (1986). Population sources are diverse, including Chandler (1987), Eggimann (1999), Kuczynski (1948-53), Maddison (2008), Manning (in process), McEvedy and Jones (1978), Mitchell (2003a, 2003b, 2003c), Modelski (2000), British Colonial Office Reports, British Colonial Office Lists, and Almanach de Gotha. General demographic sources include Mitchell (2003a, 2003b, 2003c), and country-specific statistical yearbooks. Incorporates data originally collected for Colonialism and its Legacies (co-PIs: John Gerring and James Mahoney).

3. Infrastructure information includes (a) waterways, (b) roads, (c) railroads, (d) post, (e) telegraph, (f) telephone, and (g) electricity. For each infrastructural feature, we will take note of 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 information can also be aggregated by various spatial units (empire, nation, region, city, town), depending upon the coverage.

For railroads, key sources include Bogart (2009), Models (1984), Nock (1978), Siebert (2004), Stover (1999), Westwood (2008), and Railroads and the Raj: The Economic Impact of Transportation Infrastructure, a GIS project (see Appendix C). For postal infrastructure, a key source is Union Postale Universelle Statistique Generale (1875-). A key source for a variety of infrastructure elements is a dataset compiled by Christopher Housenick (in process), British Colonial Office Reports, British Colonial Office Lists, Mitchell (2003a, 2003b, 2003c), and country-specific statistical yearbooks. Incorporates data originally collected for Colonialism and its Legacies (co-PIs: John Gerring and James Mahoney).

4. Slavery information includes measures of (a) the number of slaves taken from each ethnic group during the trans-Atlantic and Indian Ocean slave trades, as well as the locations of these ethnic groups during the 18th century and today; (b) for the trans-Atlantic slave trade, the ports of embarkation within Africa, the number of slaves embarked, the date of embarkation, and the vessel, (c) the ports of disembarkation in the Americas, the number of slaves disembarked, and the vessel, (d) the size of the slave population in the nations of the Americas in 1750 and 1830, (e) the size of the slave populations in the counties and states of the US each decade from 1790 to 1860. This incorporates data originally collected for three projects: The Long-Term Effects of Africa’s Slave Trades, Slavery, Inequality and Economic Development in the Americas, and The Slave Trade and the Origins of Mistrust in Africa (PI: Nathan Nunn), described in Appendix B.

5. Religion information includes information on Protestant and Catholic missionary activity, education, medical work, pastoral work and local congregations from the mid-19th to mid 20th centuries. Much of this data is geo-referenced so it can be connected to current countries, provinces, and cities. This section incorporates data originally collected for Project on Religion and Economic Change (PREC) (PI: Robert D. Woodberry), described in Appendix B.

6. Conflict data will be taken from various encyclopedic summaries of world conflicts. The most complete source currently is Warfare and Armed Conflicts (2002). This source alone will allow us to digitize and geocode the location of known world conflict globally since 1400. We will also collect other information, such as the number of battle deaths and the participants involved in the conflict.

These six areas, with their various sub-components, comprise the information of immediate concern to the project. Several other areas bear mention by virtue of the fact that a great deal of essentially “free” data is already available. This includes data pertaining to commerce (e.g., trade flows, trading posts, mines, crops, and arable land), language (e.g., the area in which a language is regularly practiced, the area in which it is the majority or dominant language, the total number of speakers within a jurisdiction, and density), ethnicity (e.g., the area in which an ethnic group is prevalent, the area in which it is the majority or dominant ethnic group, the total number of group members within a jurisdiction, and density), and climate (e.g., rainfall, drought, temperature).

More generally, it may be worthwhile to briefly review the data sources (beyond those already cited) available for a historical database-building project of this nature, in the immediate and long-term. The following paragraphs summarize the more detailed accounting presented in a series of appendices to this proposal. Note that wherever information is proprietary, integration with GH-GIS will involve negotiating with the copyright holders and special compensation may be required.

First, there are web and/or GIS-based projects that speak to long-term historical developments (see Appendices B and C). This includes: the Animated Atlas of African History (Brown University), the Centennia Historical Atlas (cd-ROM), the China Historical GIS Project (Harvard University), China in Time and Space (University of Washington), the Climatic Research Unit (University of East Anglia), Digital Atlas of Ancient and Medieval Civilizations (Harvard
University), the Electronic Cultural Atlas Initiative (UC Berkeley), Euratlas (Switzerland), the Mapping History Project (University of Oregon, University of Munster), the Plough (Harvard University and UCLA), the Spatial Analysis Project (Harvard University), the World Language Mapping System (Colorado Springs, CO), and various other national historical GIS projects (most of which are focused on Europe and North America [Knowles 2005]).

Second, there are cartographical sources in hard-copy format, e.g., historical atlases and antique maps. These sources are often helpful in clarifying the precise location of spatial units of importance to the project, especially those that undergo changes through time, as well as in adding new substantive information to the database. Special attention will be paid to areas of the world neglected by prior historical GIS projects so that a truly global representation of history can be created. A selection of important historical atlases is listed in Appendix D. Historical maps are available from a number of sources including AfricaMap, the David Rumsey Map Collection, the Electronic Cultural Atlas Initiative (ECAI), Gallica, Harvard University, the Library of Congress, the Mapping History Project, Wikipedia, and sources reviewed in the previous section.

Third, there are non-cartographic sources of historical information. These are essentially limitless in format and subject. However, in order to give the reader a sense for what we think the most useful of these materials are, we include a preliminary list of primary and secondary sources as Appendix E. Fortuitously, many out-of-print sources, including rare journals and archival sources, are now available on-line from Googlebooks or Open Library.

As we proceed, we will be maintaining a bibliographical database including all sources included in the project, with annotations about the topics covered and extent of coverage. This bibliographical database, which will ultimately become a searchable component of the on-line GH-GIS (perhaps using Heuristic), promises to be a boon to scholars looking for information on a wide variety of subjects in the humanities and social sciences.

III. Colonialism and its Legacies (NSF Grant SES-0648292)

This project builds on a number of extant projects, including one current NSF grant (SES-0648292), Colonialism and its Legacies: A Comprehensive Historical Dataset, directed by John Gerring and James Mahoney.

The original impetus for Colonialism and its Legacies stems from the prominence of colonialism in causal accounts of the modern world and the corresponding dearth of systematic data pertaining to the colonial experience. It has become commonplace to observe that the colonial experience shaped the modern era in profound ways. Yet, while there is general agreement that “colonialism mattered,” it is less clear what the long-term effects of this traumatic intervention actually have been. The virulence of scholarly and popular opinions about colonialism is matched by the inconclusiveness of current research. The core difficulty can be stated succinctly: scholars lack a reliable, global dataset for testing hypotheses. As a remedy for these problems, this project proposed to develop a comprehensive historical dataset focused on Anglo-European and Japanese overseas colonialism.

As part of this ongoing project, data relevant to colonial experiences around the world are in the process of being identified and collected in a single MS Access dataset. Data collection focuses primarily on large geographic entities (nations, colonies, empires) and on the modern era of European colonialism. By the end of the project period (Spring 2011), we anticipate that many of the information sources displayed in appendices B-E will have been integrated in some fashion. Currently, the project embraces more than 250 sources (primary and secondary), over eighty variables, and over 6000 political units (empires, countries, colonies, regions, and municipalities). Coverage centers on the 18th-20th centuries, with some variables and political units extending further back in time.
What this means in terms of the current project is that a good deal of the data collection outlined in section II has already been accomplished – or at the very least, begun. It remains to extend the reach of Colonialism and its Legacies and to organize this data in a historical and spatial format so that it can be easily accessed, viewed, and manipulated. Information pertaining to the British Empire, e.g., population or ethnic composition, must be understandable in terms of the changing spatial contours of the empire. Likewise for smaller units – colonies, regions, and so forth. In this sense, GH-GIS builds directly upon Colonialism and its Legacies.

IV. Dataset construction

The categorization schemes envisioned by GH-GIS will demand a good deal of time and attention. As we have emphasized, the process of integrating data pertaining to diverse subjects and from diverse sources involves not only technical issues but also decisions about which features to integrate and how to fit the new information into an overall classification. We need to decide, for example, how to define the superordinate category “religion,” as well as every subordinate category that falls within this rubric (e.g., animist, Christian, Islamic, and so forth). Every bit of information that is integrated into GH-GIS will involve a coding decision of one sort or another. Developing the taxonomy, and exercising correct coding decisions, is a massive undertaking, requiring consultation with scholarly experts (in this case, anthropologists, historians, and theologians). This will require the time of the co-PIs, graduate assistants (particularly assistants fluent in multiple language to allow translation between sources). 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.

Each observation in GH-GIS 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 (variously recorded, as described in section I), (c) attributes of interest (as in Section II above), (d) year (and more specific date, if any), (e) source (which may involve several layers of sources – as well as information about how and when it was digitized), 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 GH-GIS should be able to reconstruct the origin of each datum.

Several additional steps will be taken to deal with error, uncertainty, and incompleteness. First, PDF copies of all original sources will be preserved, and (if possible) linked directly to the dataset so that users may reference the original whenever issues of fact or interpretation arise. Second, all data will be collected and retained in a format that hews closely to the context of the primary source, as indicated. 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.

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 also be re-coded so as to render them commensurate with similar information drawn from different eras and areas. Indeed, many of the coding categories listed in Section II 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,” as well as the hotly contested political variables (sovereignty, colony, et al.). The categories listed in Section II represent an attempt to create universally applicable concepts that 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 (through folksonomies).

To complement the primary features described in Section II, secondary categories will be prepared in order to present the material in a fashion accessible to 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 linear combination). 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 basic-level categories contained in Sections I and II are highly flexible, and may be aggregated in multiple ways so as to address larger issues of historical and contemporary concern.

Let us return for a moment to the twin problems of reliability and comprehensiveness. We recognize that information necessary to populate the gazetteer and to code the numerous topics identified in sections I and II is not ubiquitous and, where available, may not be terribly precise. In general, we anticipate that the quantity and quality of information will be highest for Europe, Japan, and China. The colonial holdings of England, Germany, the Netherlands, and Japan are blessed with more historical information than the colonial holdings of other powers. In short, some parts of the world are more knowable than others (and this of course varies by historical era). In all regions of the world, the quality and quantity of information generally diminishes as one moves back in time: the world in the nineteenth century is more knowable than the world in the sixteenth century – not to mention pre-historical eras. Finally, 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.

Our aim is not to create an immaculate dataset – free from error, perfectly precise, and comprehensive. It is, rather, to reach a level of accuracy, precision, and comprehensiveness that is comparable to that displayed in the work of specialists on these subjects. GH-GIS represents a compilation of existing knowledge about the past; as such, it must suffer the same shortcomings that extant historical knowledge presents. However, insofar as traditional historical research has illuminated parts of the human past – insofar, that is, as we know anything at all about human history – GH-GIS will offer a major advance, for it promises to unify that knowledge, to make it more accessible, and to provide mechanisms for estimating error in a systematic fashion (e.g., by comparing estimates drawn from different sources).

V. Infrastructure and Applications

The software architecture of GH-GIS will provide a web-based system for querying, displaying, storing, and collaboratively editing historic spatial information. Our proposed system builds on the

---

5 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.
Harvard University Infrastructure for Geospatial Collaboration (HUG), recently developed for the AfricaMap project (http://africamap.harvard.edu) by the Center for Geographic Analysis (CGA) at Harvard. HUG represents a framework for organizing spatial and non-spatial data, making it findable, map-able, comparable, and downloadable from a web browser. Data in HUG is served live from various systems inside and outside of Harvard through the internet, brought together to form an integrated map image for the user. In addition, the data that are stored on the GH-GIS servers are made available to other applications as map services. In this way HUG is becoming a node on the emerging geo-web.

The system enables multiple organizations to collaboratively build a variety of historical atlases at a variety of scales, using the best historical map source materials available. Each collaborating organization may store and manage their data locally, following standard approaches for sharing metadata and map data, thus enabling GH-GIS users to find and access their local data holdings within the GH-GIS system. This architecture makes it possible to store and access a virtually unlimited volume of historical spatial information, source and derived. The data stored at Harvard will be made permanently and publicly available as it will be stored in endowed systems such as the Harvard Geospatial Library (http://hul.harvard.edu/ois/systems/hgl/) and the DataVerse Network (http://thedata.org/).

The HUG infrastructure provides several essential components for a collaborative global historical GIS. These include:

1) A web-based mapping client which supports the following user functions:
   a. Ability to turn layers on and off, and to 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 earth and terrain imagery, these services are unmatched.
   g. Ability to view layers in Google Earth in addition to the web client.
   h. Bookmarking - encapsulation in a 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 also 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 fast display and high scalability.
   c. A service oriented architecture which makes content available to other systems as web services and is also able to consume services which other organizations make available.
   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 20 layers composed of 3,163 map sheets for the years 1612 to 1990, while the Portsmouth team will provide 1:500,000 scale historical mapping for most of the world.

Additional enhancements to HUG will be necessary to support the full requirements of a global, historical GIS. The enhancements proposed here will allow for organizations and researchers to collaboratively build datasets derived from source materials such as scanned historic maps. Derived datasets may include historic layers such as gazetteer, trade routes, political boundaries, transportation routes, etc. This system will be built upon the successful model of OpenStreetMap (http://wiki.openstreetmap.org/) and will make use of the OpenStreetMap source code (which is open source) 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.

Organizations such as map libraries can use the system to scan, georeference, and make available their historical maps for use by those interested in building derived historical GIS layers. They will also be able to add new locations to the gazetteer by extracting places from these historical maps. Gazetteer entries may include not only points but also linear and polygonal features such as rivers, roads, or administrative boundaries. One location may have multiple place names through history. The system will provide a robust time model to support time range information (begin date, end date, and level of certainty for begin and end dates) for every feature in the system. For features without identifiable date ranges, a text field will be populated with any meaningful time information. Even though such time information in a text field cannot be searched and rendered by a time range, it is indispensable for historical research. The system will also support tagging of features. A tagging ability allows users to organize materials according to their own needs. When many people use tags within an online community, tags can provide a dynamic and flexible alternative to predefined ontologies, i.e., folksonomies. (Robust folksonomies can be found in communities such as flickr and delicious.)

We propose to build the following six tools as enhancements to the existing HUG architecture, adapted for use in the GH-GIS project:

User Login Tool (account manager): We plan to create this tool to delimit and track user access through a user ID/password to administer who can see, modify, or upload materials. This allows for both control of data and more specialized engagements. This will also provide the means to set up shared work spaces (virtual seminar rooms) for groups of scholars or faculty and students working on a particular topic to assemble, analyze, discuss, create, and submit new documents.

User Upload Tool: This will enable uploading of both individual data and bulk content, whether the materials are scanned images, GIS data layers or tabular records; already geo-referenced or needing online georeferencing within the system. The tool will be designed to capture user-specified source data, metadata, georeferencing and time definitions, default display symbology (such as size of the points or color of the shaded areas), legend, searchable attributes, access permission, and other information. Once captured, this information will be entered into the corresponding parts of the geodatabase.

User Georeferencing Tool: The traditional process of geo-referencing maps and other data is time-consuming, tedious, and often-difficult. In light of the enormous potential of GIS this is unfortunate. We plan to create a tool that will simplify this process, and allow even non-specialist GIS users to do their own geo-referencing, starting with country, place name, district, or culture/ethnicity as location references. Collection owners will identify fields in their metadata that are likely to contain place name or ethnicity references. Once this is done, they will be able to invoke the automated georeferencing function, which will check place names for the defined fields against the gazetteer to
determine preliminary lat/long coordinates. The owner will then be able to check and correct the coordinates assigned using an interactive mapping environment. Coordinate values will be stored in the database in the OGC standard Well Known Text format which can be output to any spatial format. The projection space used for storing coordinates in the system is Geographic (lat/long) and the datum used will be WGS 84.

The carefully defined best practices for georeferencing provided by the BioGeomancer project will be used as a starting point and modified as necessary. Although the BioGeomancer project is oriented toward the georeferencing of natural science specimens, most of the core ideas hold for referencing other type of materials. The key elements for success are good source information and good maps.

Users will also be able to georeference scanned map images. The user will upload an image, then using existing georeferenced maps as references, identify tie points between the uploaded image and existing materials. Once enough tie points are defined, the system can reference the scanned map to geographic space. The system will allow users to perform this task as many times as desired to obtain a good fit between the map and the “real” world. Users will be able to comment on one another’s referencing work and even redo referencing done by others. The object owner will then have the option of choosing whether to accept a corrected version.

**Scholarly Exchange Tool and Syndication:** We plan to create a tool that will enable logged-in users to add map markup with comments to maps and other data in GH-GIS. This tool will provide a modern version of historic museum collection object cards, in which area specialists corrected past information and contributed to new data. This tool will also serve as a means to document and credit contributors of new data layers. The scholarly exchange tool will enable scholars from various disciplines to have conversations about data on GH-GIS.

Users will be able to comment on each other’s maps by placing point, line, or polygonal features on the map and adding comments to them. The system will automatically track the user, date, location, and map reference for each comment. The user will be able to email their map markup to colleagues. A URL will be automatically generated and included in the email along with the comment. Upon clicking on the URL, the email recipient will be taken into GH-GIS and the comment will be displayed in the context of any maps that were displayed when the email sender created the comment.

Any comment created by a user may be syndicated using the syndication standard for blogs, RSS. Users will be able to choose to make any comment appear in their RSS feed which anyone will be able to subscribe to. This function will make comments made in GH-GIS usable by any blog in the world. As with the emailed feed, a subscriber reading a feed will have a URL which will take them directly into the GH-GIS system to view the comment within the map context in which it was made. We envision that syndication will be an important means of dissemination and will bring new users into the site where they will be positioned to continue a discussion by adding their own comments.

Links which point to content on other servers are subject to breakage when the linked-to site changes the URL or disappears. To address this we will develop an administrative tool that can be set to periodically check all links in the system on a regular basis and generate a report for the administrator. After a defined number of failed attempts to access a given link, the system will email the creator of the link to notify her of the problem.

**Time Delimiting Tool:** GIS is an important tool for displaying, searching, and comparing geo-spatial data. The concomitant complexities of time-delimited data (layers with date attributes) have posed certain challenges in terms of accessibility (searches) and display within a GIS format. We plan to create a tool which will provide the option for allowing machine readable time (dates and/or date ranges) to be created and saved, and then searched and displayed.
Users will be able to precisely define time ranges (start date and end date) for any feature created in the system. The international standard for time and date representation, ISO 8601, will be used to store date ranges in the database. ISO 8601 is compatible with the Dublin Core metadata specification which is used by OAI and is also compatible with the KML standard for geographic representations used by Google Earth.

GH-GIS will support the visualization of time enabled spatial materials by outputting a version of the features in KML format for visualization using the Google Earth time bar. In addition, the GH-GIS query interface will support searches using start date and end date, and the GH-GIS Time Graph view will present materials stored in GH-GIS which have time ranges defined.

**Search and Visualization Tool:** Users will be able to search for materials in GH-GIS using the traditional filters of Title, Author/Creator, Subject, Language, and Resource Type. Users also will be able to search by time range, choosing a “time window” filter for their search.

In addition to the standard query result list, three interactive graphic views will be available to help the user characterize a collection or the results of a query: a Map view; a Time Graph view; and a Word Cloud view. The Map view will display a heat map showing the distribution of materials across the geographic extent of the selected set of materials. Areas with greater concentrations will show up in a different color. The Time Graph will display a graph of the materials distributed across the time frame they represent. The Word Cloud will use the approach of a standard tag cloud except that it will be possible to choose fields such as Language and Author to create focused clouds in addition to the more traditional unstructured tag clouds.

The relationship between the four pieces (list, map, time graph, word cloud) will be dynamic. Zooming in on the map to show a subset of the results on the map will filter the rest of the views (time and tag) which will change to reflect the new spatial filter. Similarly, changing the time range will filter the other 3 views, and selecting a word from the word cloud will cause the other views to adapt. In this way a user will be able to explore large collections of digital materials with a better a priori sense of the distribution of materials.

Presently, the HUG platform has accomplished the first phase of its planned development, relying on internal funds from Harvard University. As of the end of June 2009, these contributions total roughly $300,000. Additional development of this architecture depends upon external funding.

In order to adapt HUG for the additional data loads and functions envisioned for GH-GIS several capabilities will need to be developed. These may be summarized as follows: (1) tools for access control at user level and layer level, (2) tools to support distributed data storage and centralized searching and display, (3) tools for data layer and metadata submission, (4) tools to support preliminary georeferencing of online digital collections, (5) tools for time delimiting, and (6) tools for redlining, tagging, editing and exchanging comments. Further details on the proposed architecture of GH-GIS can be found in Appendix A.

**VI. Implementation and Sustainability**

In order to be successful, GH-GIS must be a collaborative effort, involving scholars, research centers, libraries, and existing online systems that currently provide information about countries and regions. Our hope is to facilitate a wide-ranging dialogue in which scholars can discuss their experiences, their needs, and their ideas for the future development of GH-GIS.

It is vital, in particular, to gain agreement on a format for information storage and retrieval that spans the software practices and serves the needs of multiple end-users. Equally important, we must engender a sense of shared ownership over the enterprise. If GH-GIS is to provide a public good for the scholarly community it is important that it be embraced by the community. This requires continual communication and coordination, and thus serves as an essential task of the GH-GIS
project. In the course of this discussion, we expect that many of the instruments and some of the objectives identified here may be modified.

Our goal is to allow all aspects of GH-GIS to be freely available to all users, without preconditions. This includes the download of all data files (including GIS shape files), which is easily imported for use with statistical programs (e.g., R, SAS, Stata). It also includes the technical architecture of GH-GIS, which will be open source.

Evidently, the utility of GH-GIS is proportional to the accuracy of the included data, the range of topics encompassed by the data, and the completeness of coverage (through time and through space). The project, as sketched here, is a starting-point, not a final resting place. It is our intention to provide enough value-added that the project will stimulate scholars working around the world on diverse subjects to correct errors (or at least comment on possible errors) and to contribute original data. If this happens, the project will grow organically in wiki fashion.

This brings us to the problem of sustainability. A project of this scope must be maintained, revised, and expanded if it is to continue to serve its myriad functions for its myriad clienteles. As we see it, five issues of sustainability must be faced.

First, there is the need for permanent archiving and versioning of digital materials. With respect to maps, we have proposed to archive both raster and vector data in the Harvard Geospatial Library. With respect to quantitative datasets with spatial attributes, we propose to archive and version these through Harvard’s Dataverse Network.

Second, one must be concerned with ongoing data acquisition of and cleaning/processing of contributed data. There is a growing interest in setting up processes for “publishing” peer-reviewed datasets and databases, which will hopefully encourage scholars to share datasets created in the course of their research. GH-GIS would provide a venue for such datasets and would assist in the establishment of standards for scholars who wish (quite naturally) to be credited for data created in the course of their research.

Third, there is the technical task of maintaining the online system and maintaining interoperability with other systems over time. This is a concern that must be taken up by the (growing, but still rather small) community of scholars who work with GIS-based historical materials.

Fourth, there is the question of gatekeeping, specifically, how to encourage and to vet new contributions to GH-GIS. We see no reason to restrict contributions to GH-GIS. However, we see a strong need to distinguish between those that have been vetted by experts and those that have not. To this end, we envision a board of directors that would govern a highly decentralized structure of experts in different areas encompassed by GH-GIS. Areas would be organized by topic (e.g., ethnicities, roads), by area (e.g., Middle East), and by historical era (e.g., contemporary, ancient, prehistory). Accordingly, the job of the “ethnicity” expert would be to manage this layer (or set of layers) of GH-GIS, including canvassing for new data sources, identifying those deemed most authoritative, and vetting in-coming contributions. It is our hope that this task would be an enjoyable one for the appointed “experts” and would be carried out in conversation with other (un-appointed) experts in each area. Note that the chosen experts are presumably individuals who are already engaged in this activity; the only added burden (or perhaps benefit) is the use of software developed for GH-GIS.6

And finally, there is the question of how to pay for the staff time necessary to achieve the foregoing tasks. In the long term, sustainability will require that libraries see it as a necessary part of their information licenses. In the best of all possible worlds, major research universities around the world would, through their libraries and GIS centers, implement the system and oversee local uploading of datasets, making it freely available to the public. What is done at one institution would

---

6 Useful discussion of the open-source model can be found in Weber (2005).
become visible to all others (at least at the metadata level for licensed materials). This is the principle on which the Dataverse Network, the most promising means of registering and storing datasets, operates. Whether this would work in practice is an open question. Ultimately it may be necessary to follow the JSTOR model, charging fees scaled appropriately to the financial circumstances of the licensing institutions. However, these choices will only face the scholarly community when the viability and potential of the system is demonstrated in practice, thus this application.

VII. References


“European State Finance Database.” [Comprises dozens of datasets housed at the Arts and Humanities Data Service (AHDS), including revenue and expenditure data for countries and cities, mostly though not exclusively in Europe, back to the Middle Ages.]

Great Britain Colonial Office Reports.


Hentschel, Paul. [various years]. “ICOW [International Correlates of War] Colonial History Data Set.” http://garnet.acns.fsu.edu/~phensel/

Hertslet, E. 1875. *The map of Europe by treaty: showing the various political and territorial changes which have taken place since the general peace of 1814*. London: HM Stationery Office.

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.”


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


Timelines Thesaurus, developed by The National Monuments Record (part of English Heritage).
www.taxonomywarehouse.com/vocabdetails_include.asp?vVocID=240
Appendix A:

Project Architecture

Current AfricaMap (HUG) System Characteristics

**Web-based** - The system takes advantage of the latest techniques for making large amounts of data and mapping discoverable and usable through a standard web browser.

**Public access to holdings** - Core holdings will be put in the public domain or licensed using a Creative Commons type license wherever possible. This means that researchers anywhere in the world will be able to download and use these original materials without major restriction.

**Base mapping** – Historic base maps for Africa are developed by scanning, cropping, and georeferencing maps from the Harvard Map Collection and elsewhere. Maps are digitized at a range of scales and for a range of time periods.

**Dynamic gazetteer** – The gazetteer together with the base map form the core of the AfricaMap system. These two datasets support one another over time, allowing the gazetteer to grow and improve, which will make it easier to find places on the base map.

**Collaborative approach** – Some tools to support collaboration between researchers are provided. In the first version a permalink feature will allow any view of the system to be captured in a URL which can be shared. In the next phase user created maps and map markup tools are anticipated. Researchers will be able to download base mapping and other datasets.

**Multiple scales** - The system will support research at a variety of scales from sites or cities to country or continent-wide projects.

**Multiple media types** - The system will support access to many types of media in addition to spatial data, including photos, maps, text, video, audio, and KML for Google Earth display.

**Long term data access** - Once maps are scanned, digitized, georeferenced it should not be necessary for anyone in the world to repeat that work. Making digital materials available over time is not easy because technology changes. Techniques will be used to ensure long term access to public domain digital materials wherever possible.

**Improves over time** - The goal is to enrich the collection over time by sharing with other libraries and collections. Users will be able to submit data to Harvard using an online form.

**Usability** - Ease of use is of primary importance. It must be easy and quick for non-technical people to find the information they need. Researchers are the end users of this system and will be consulted frequently to guide the design of the user interface.

**Text-based search of contents** - Google-type text search against the contents of the entire system is possible with results displayed on the map.

**Interdisciplinary approach** - The system will bring together mapped data (and facilitate the mapping of data) from a wide range of disciplines including archaeology, public health, history, linguistics, literature, zoology, natural resources to name a few.
**Scalable** – The data in the system will be cached as it is used. This approach greatly increases performance and reduces server load, making the system far more scalable than a traditional web-GIS.

**Services oriented architecture (SOA)** - The system will support access by other web and desktop systems and will in turn be able to access and display the maps on AfricaMap directly via web services. This means that other systems will not have to download the data to access it within their applications.

**Cross Platform** – HUG can serve data services to other types of GIS platforms including ArcMap desktop and ArcGIS Server. In addition, HUG can display data served up from other platforms. Data formats used will be open specification ones such as GeoTIFF, JPG2000, KML, and Shape.

**Open Source** – The software that runs AfricaMap is Open Source and available for users and organizations inside and outside Harvard to obtain and build upon.

**Current AfricaMap (HUG) Architecture**

This diagram presents the architecture of the current AfricaMap system.
AfricaMap uses a Javascript-based map client to display map tiles from a local tile cache as well as tiles from remote servers such as Google’s. All map data is rendered as tile caches including the gazetteer (place name database), political boundaries, historic maps. Tile generation occurs based on WMS (Open Geospatial Consortium Web Map Service) requests.

Caches are generated from the original GIS data by map renderers MapServer and GeoServer. All queries against vector (point/line/polygon) GIS data datasets are handled by PostGIS, a spatial relational database. Tiles are stored in an open tile format called TMS (Tile Map Service).

This architecture makes it possible for AfricaMap to connect to remote WMS instances and cache them for display in AfricaMap. Alternatively AfricaMap can connect directly to a remote WMS for...
display in AfricaMap. In addition, any data layers in AfricaMap can be made available to other organizations via WMS or TMS.

Data can be stored in any source projection and reprojected on-the-fly to Spherical Mercator for overlay with commercial map services such as Google Maps.

Currently Google Earth versions of some of the raster data are stored on disk, though this could eventually be handled by reading directly from the tile caches.

**Proposed GH-GIS Architecture**

This diagram presents the existing architecture in black and the proposed in blue. Highlighted are areas of significant software development.

![Proposed GH-GIS Architecture Diagram](image)

**Figure B.2 - Proposed GH-GIS Architecture**

The proposed architecture borrows from AfricaMap and adds some powerful new features to support ongoing collaborative historical GIS datasets development between geographically distributed participants.

An important enhancement is the creation of a mechanism for leveraging the wealth of historic geographic information in the world’s paper map collections by making it easier for organizations to publish their maps in a useful way. This approach uses tile caches as in AfricaMap but with some twists. First, the list of available tile caches is put on the web in the form of a database (GIS Catalog Server) which can be updated by those who wish to contribute new caches to it, and which can be queried by users searching for maps. These caches do not need to reside on a central server, but can
reside anywhere, including on the server of the organization that owns the original paper map. No special software is required in order for the organization to serve them efficiently. The caches just need to be stored in web accessible space.

The GH-GIS project proposes the development of an open source tool for these organizations to use to create tile caches from their georeferenced maps and register metadata for the caches along with the cache location URL on the central Catalog Server. The Catalog Server will make its contents searchable via an open API so that others will be able to build a client that can query the catalog and retrieve tiles from wherever the tiles reside.

It will also be possible for those with historical GIS data of any type to register their data on the Catalog Server. They will have the option of inputting a WMS URL to the Catalog Server or uploading the data.

In addition to a framework for distributed sharing of historic map data, tools will be provided, (HUG Python Application Code in diagram), for collaboratively editing historic GIS layers based on the tile cache backdrops. Tools will allow users to add features to existing layers with defined schemas. A key layer will be the historic gazetteer (place name database). This layer will evolve and improve as more and more historical maps are uncovered and more place name information is manually extracted from them.

An important feature of the editing tool will be provenance maintenance. The reference information for the base map used to determine the location of a place name in the gazetteer will be permanently linked to the derived place name feature.

In addition to the manual extraction of place name features from historic maps, we propose the development of an innovative means of gazetteer population: a specialized version of OCR (optical character recognition) which is adapted to support the extraction of place names and locations (lat long) from scanned historic maps. We are calling this technology Geo-OCR. Such an approach, in a collaborative environment such as this does not have to be perfect to be useful. This is discussed in more detail in “Paper Map Publishing Data Flows” diagram.

A global search ability will be provided in GH-GIS. It will be possible to search the entire system (all historic GIS layers and all caches) for a key word and display the results highlighted on the map. Clicking on highlighted features on the map returns information about any overlapping features which match the text search.

**High Level View – Stages in Collaboration**

This diagram shows the source data providers on the bottom and the map viewing and editing clients on top.
Step 1 – Librarian at organization with paper map collection scans map and (optionally) georeferences and (optionally) extracts place name features, creates tile cache, puts cache on web accessible space. An online tool for georeferencing scanned maps will be provided.

Step 2 – Librarian notifies the Catalog Server of the existence of the tile cache by URL, fills in basic metadata regarding source, date, georeferencing technique used.

Step 3 – Researcher looking to find historic maps to support work in a given region, queries the Catalog server using the GH-GIS web client, and finds list of matching datasets.

Step 4 – Researcher turns on desired tile cache and request is made directly to the tile cache server.

Step 5 – The researcher uses the tile cache as a backdrop and digitizes new place name features to add to an historic layer.
Paper Map Publishing Data Flows

There are several ways in which raw historical map images can be made available to the GH-GIS system for use in building derived historical GIS datasets.

Figure B.4 - Paper Map Publishing Data Flows
Appendix B:
Related Projects Undertaken by the Project Team

AfricaMap
URL: http://africamap.harvard.edu
Location: Center for Geographic Analysis (CGA), Harvard University
Principals:
Ben Lewis, Center for Geographic Analysis, Harvard University
Suzanne Blier, Fine Arts, African & African American Studies, Harvard University
Peter Bol, East Asian Languages & Civilizations, CGA, Harvard University
Wendy Guan, Center for Geographic Analysis, Harvard University
Status: Ongoing.
Period: Mostly contemporary.
Availability: Most of the data will be made public.
Description: AfricaMap sets out to address the problem of data availability for Africa. At its core is a digital base map of the continent, viewable dynamically at a range of scales, and composed of the best cartographic mapping available. Behind the scenes a gazetteer starting with over 1 million place names provides rapid navigation to specific locations on a vast landscape. AfricaMap stores and references data from all disciplines including the humanities, the social sciences, and certain of the natural sciences (e.g., geography). Four broad categories of mapping are included: 1) contemporary maps available in the Harvard Map Collection and elsewhere; 2) data on African cultures (e.g., population density, languages, subsistence, beliefs, technologies, forms of shelter) provided by the Human Relations Area Files; 3) historical maps of the continent (geo-referenced); and 4) scholarly maps focused on Africa in various historical periods. The project also forms a framework for referencing an array of materials important for different areas and periods, e.g., photographs, charts, videos, music segments, and spoken language exemplars.

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, East Asian Languages and Civilizations, Center for Geographic Analysis (CGA), Harvard University.
Merrick Lex Berman, Center for Geographic Analysis (CGA), Harvard University.
Man Zhimin, Director of the Center for Historical Geography, Fudan University, Shanghai.
Collaborators:
Lawrence Crissman, Department of International Business & Asian Studies, Griffith University.
Jianxiong Ge, Center for Historical Geography, Fudan University, Shanghai.
Mark Henderson, University of California, Berkeley.
Xiaofeng Tang, Historical Geography Research Institute, Peking University.

**Period:** 222 BCE-1911.
**Area:** China.
**Status:** Ongoing.
**Availability:** Free.

**Description:** 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.

**Colonialism and its Legacies**

**Location:** Boston University and Northwestern University
**Sponsors:** National Science Foundation
**Principals:**
  - John Gerring, Department of Political Science, Boston University
  - James Mahoney, Departments of Sociology and Political Science, Northwestern University

**Collaborators:**
  - Paul Barclay, Department of History, Lafayette College
  - Neil Englehart, Department of Political Science, Bowling Green State University
  - Jonathan Krieckhaus, Department of Political Science, University of Missouri–Columbia
  - Charles Kurzman, Department of Sociology, University of North Carolina
  - Patrick Manning, Department of History, University of Pittsburgh
  - Erzen Oncel, Department of Political Science, Boston University
  - Claudia Rei, Department of Economics, Vanderbilt University
  - Jennifer Rosen, Department of Sociology, Northwestern University
  - Nicolas van de Walle, Department of Government, Cornell University
  - Robert D. Woodberry, Department of Sociology, University of Texas at Austin

**Period:** 15th century-present
**Area:** Global.
**Status:** Ongoing.
**Availability:** All of the data will be made public on completion of the project.

**Description:** A comprehensive historical dataset focused on Anglo-European and Japanese overseas colonialism since the fifteenth century. Using the country-year or country-decade as our unit of analysis, this project attempts to code all cases from 1400 to the present along a variety of dimensions related to colonialism (specifically) and political, social, cultural, and economic development (more generally).

**Digital Atlas of Ancient and Medieval Civilizations (DAAMC)**

**URL:** medievalmap.harvard.edu [not yet public]
**Location:** Center for Geographic Analysis (CGA), Harvard University.
Sponsor(s): the Fund for Instructional Technology of the Provost of Harvard University, McCormick’s Distinguished Achievement Award from the Andrew W. Mellon Foundation.

Principals:
- Michael McCormick, Department of History, Harvard University.
- Guoping Huang, Center for Geographic Analysis (CGA), Harvard University.

Period: Classical and Medieval eras.

Area: Middle East, Europe.

Availability: Public, upon completion.

Status: Ongoing.

Description: The first high quality digital maps of the Roman Empire and Medieval Europe suitable for continuous improvement and adaptation in teaching and research worldwide. Interactive maps (GIS digitized) will allow users to zoom in/out, click on a site and see related information about this place, and even listen to instructor's tour of the monument, site, or excavation, as well as see all digital resources, esp. photos. This project uses geodatabases as the central container to host all the spatially enabled historical information, then builds applications to serve the information out. The web-based DAAMC is constructed with ARCMAP 9.2 and consists of a substantial and growing number of geodatabases documenting multiple aspects of ancient and medieval civilization in their spatial dimension, including ancient and medieval towns, Roman roads, medieval monasteries, ecclesiastical provinces, the diocesan boundaries of medieval France, Crusader routes, Roman and medieval shipwrecks, epidemics, medieval universities, the rise and fall of kingdoms and empires, etc. It allows an unparalleled array of innovative spatial and temporal analyses of all aspects of the civilizations of western Eurasia in the first 1500 years of our era, as well, of course as the generation of original maps of differing aspects of ancient and medieval civilization.

Great Britain Historical Geographical Information System (GB-HGIS)

URL: www.port.ac.uk/research/gbhgis/abouthistoricalgis/

Vision of Britain Through Time: www.visionofbritain.org.uk/index.jsp

Location: University of Portsmouth

Sponsor(s): Joint Information Systems Committee, European Union FP6, Big Lottery Fund, Economic and Social Research Council, Leverhulme Trust, Wellcome Trust, and others

Principal: Humphrey Southall

Period: Mainly 1801-2001

Area: Great Britain

Availability: Free for academic use, plus public on-line site A Vision of Britain through Time

Status: Ongoing.

Description: The original Great Britain Historical GIS (1994-2000) held computerised boundaries for the main statistical reporting units of late 19th and 20th century Britain, linked to a large database of historical statistics from the census, vital registration and other sources. Since 2001, the system has been rebuilt around a novel ontology-based framework, in which both polygons and statistics are held as attributes of "units", which can be anything from a parish to the United Kingdom as a whole. This architecture can also hold any number of variant names for each unit, and the system was designed to work not just as a statistical mapping system but as an archival name authority, and to support a popular web site for local historians. Also with these aims, the contents have been broadened to include a large collection of scanned and geo-referenced historic maps, and a great deal of "geo-referenced text", including the full text of three late-19th century descriptive gazetteers and the largest on-line collection anywhere of historical British travel writing. Recent work has also
included developing a prototype European system, with detailed coverage of Estonia and Sweden, as well as Britain, plus outline coverage of the rest of Europe. The system is designed to support many data standards, including Dublin Core, the Alexandria Digital Library Gazetteer Content Standard, the Text Encoding Initiative guidelines, and the Data Documentation Initiative standard.

Mapping Europe's Historic Boundaries and Borders
URL: http://www.geog.port.ac.uk/hist-bound/
Location:
Sponsor(s):  
Principal(s): Michael Goerke; Humphrey Southall
Period:
Area:
Availability:
Status:
Description: Conference held in Florence/Firenze, Italy in June 2000.

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.

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.  
**Status:** Working Paper.  
**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 D. Woodberry, Department of Sociology, University of Texas at Austin.  
Juan Carlos Esparza, Department of Sociology, University of Texas at Austin.  
**Period:** 1813-present  
**Area:** Non-western.  
**Availability:** Free, once completed.  
**Status:** Ongoing.  
**Description:** 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. 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. Our data sets, based on missionary, church, and census documents, permit an unprecedented look at the cumulative effect of religious groups on economies over 190 years and in every non-Western society. Data collected through this project provides researchers with an unparalleled new resource to help answer questions regarding the health, educational, legal, religious, and economic conditions of many nations across nearly two centuries. PREC has compiled data on virtually all Protestant missionary activity, education, and medical work from 1813-1968, most Catholic religious, educational, and medical work from 1882-the present, and created digital maps of Protestant mission stations from 1813-1923 and are working on maps of Catholic ecclesiastical jurisdictions from 1444-the present. We have spatially linked these data to current national and provincial boundaries, and extensive geo-climatic information, and are working on linking them to Latin American census data from 1900 – the present.

**Spatial Analysis Project**  
**URL:** TBA  
**Location:** Fairbank Center for Chinese Studies, Harvard University.  
**Sponsor:**  
**Principals:** G. William Skinner (deceased), Peter Bol.  
**Period:** 19th-20th centuries.  
**Area:** China, Japan, France.  
**Availability:** Publication will begin in Summer, 2009.  
**Status:** Ongoing.  
**Description:** Datasets and GIS files for 19th c. China, Japan, and France, and late 20th c. China created by the late G. William Skinner.
Appendix C:
Related Projects Undertaken Elsewhere

Alexandria Digital Library (ADL)
URL: www.alexandria.ucsb.edu/
Location: Davidson Library's Map & Imagery Lab (MIL), University of California at Santa Barbara.
Sponsor(s):
Principal(s):
Period: Unbounded.
Area: Global.
Availability: Unclear.
Status: Ongoing.
Description: [From the web site…] A distributed digital library with collections of georeferenced materials. ADL includes the operational library, with various nodes and collections, and the research program through which digital library architectures, gazetteer applications, educational applications, and software components are modeled, prototyped, and evaluated. ADL provides HTML clients to access its collections and gazetteer, and provides specific information management tools, such as the Feature Type Thesaurus for classing types of geographic features, as well as downloadable software code. The MIL team is working with development teams throughout the world to establish remote, independent, yet federated ADL nodes featuring local geographically referenced collections.

Climatic Research Unit (CRU)
URL: www.cru.uea.ac.uk/cru/data/
Location: University of East Anglia
Sponsor(s): University of East Anglia
Principal(s):
Period: Varies by topic.
Area: Global.
Availability: Posted on-line.
Status: Ongoing.
Description: [From the web site…] “The Climatic Research Unit is widely recognised as one of the world’s leading institutions concerned with the study of natural and anthropogenic climate change. Consisting of a staff of around thirty research scientists and students, the Unit has developed a number of the datasets widely used in climate research, as well as statistical software packages and climate models. The aim of the Climatic Research Unit is to improve scientific understanding in three areas: (a) past climate history and its impact on humanity; (b) the course and causes of climate change during the present century; (c) prospects for the future. The Unit undertakes both pure and applied research, sponsored almost entirely by external contracts and grant from academic funding councils, government departments, intergovernmental agencies, charitable foundations, non-governmental organisations, commerce and industry.” Data are available for various climate features, including rainfall, drought, and temperature. Some of this is very fine-grained, e.g., temperature and rainfall on a monthly basis back to 1900, with a resolution of .5 degree by .5 degree.

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

European Historical GIS Initiative
URL: http://www.hgis.org.uk/ehgis.htm
Location:
Sponsor(s):
Principal(s): Ian Gregory.
Period:
Area:
Status: In process.
Availability:
Description: [From the web site...] A consortium of academics from Lancaster University (UK), the Queens University Belfast (UK), the Institute of European History, Mainz (Germany) and Indiana University Purdue University Indianapolis (USA) have been exploring the potential to develop an integrated European Historical GIS. To this end, a number of meetings have taken place, most recently at the Queen's University of Belfast (Feb. 2005) and the University of Mainz (Feb. 2007). A further meeting took place at Lancaster University in December 2007 and we held a conference session on this theme at the European Social Science History Conference in Lisbon at the end of February 2008. As an initial part of this programme we are conducting an inventory of Historical GIS resources that currently exist in Europe or that are relevant to European history.

Ethnologue
URL: http://www.ethnologue.com/info.asp#contact
Location:
Sponsor(s):
Principal(s):
Period:
Area: Global.
Availability: Posted on-line but proprietary.
Status:
Description: Web and print versions are identical. [From the web site...] Features: 6,912 language descriptions organized by continent and country; 39,491 primary names, alternate names, and dialect names; 208 color maps showing location and distribution of languages; Unique three-letter
identifier for each language from the International Standard ISO 639-3; Statistical summaries with numerical tabulations of living languages and number of speakers by continent, by language size, by language family, and by country

**Gallica**

URL: http://gallica.bnf.fr/
Location: Bibliotheque Nationale de France.
Sponsor(s): Bibliotheque 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 Historical Croplands Cover dataset**

URL: www.sage.wisc.edu/download/crop1700/hist_croplands.html
Location: Nelson Institute for Environmental Studies, University of Wisconsin-Madison.
Sponsor(s):
Principal(s): N. Ramankutty; J.A. Foley.
Area: Global.
Availability: Free, available for download.
Status: Completed.
Publications:
Description:

**Global Land Precipitation Dataset**

URL: www.cru.uea.ac.uk/~mikeh/datasets/global/
Locations: Climate Research Unit, University of East Anglia
Sponsor:
Principal: 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.

Human Relations Area Files (HRAF)
URL: http://www.yale.edu/hraf/
Location: Yale University, New Haven, CT
Sponsors:
California State University, Fullerton
University of Colorado, Boulder
George Mason University, Fairfax, Virginia
University of Florida, Gainesville
Harvard University, Cambridge, Massachusetts
Indiana University, Bloomington
University of Iowa, Iowa City
University of Kent, Canterbury, United Kingdom
Kyoto University, Japan
University of Michigan, Ann Arbor
National Museum of Ethnology, Osaka, Japan
State University of New York, at Buffalo
Northwestern University, Evanston, Illinois
University of Pennsylvania, Philadelphia
University of Pittsburgh, Pennsylvania
Seoul National University, Seoul, Korea
University of Southern California, Los Angeles
Texas A&M University, College Station
University of Utah, Salt Lake City
Yale University, New Haven, Connecticut
Principals:
Carol Ember, Yale University
Melvin Ember, Yale University
Period: 16th century – present.
**Area:** Global.

**Availability:** HRAF will make databases available to the public that link coded data on the cultures in HRAF to GIS information. These datasets will be usable by anyone. HRAF continues to rely on membership dues to support its activities. Therefore access to eHRAF World Cultures will be limited to members of the HRAF consortium. With special dues for consortia, and even lower dues for Native American, First Nations and countries with less developed economies, eHRAF World Cultures is now much more affordable than before. HRAF has not raised its dues in over 10 years.

**Status:** Ongoing.

**Description:** HRAF is an internationally recognized organization in the field of cultural anthropology. Founded in 1949 at Yale University, HRAF is a not-for-profit membership consortium of universities, colleges, and research institutions. Its mission is to encourage and facilitate worldwide comparative studies of human behavior, society and culture. In the 1930s, behavioral scientists at Yale's Institute of Human Relations started to develop classification systems of cultures ([Outline of World Cultures](#)) and of subjects ([Outline of Cultural Materials](#)) to be used to rapidly retrieve information relating to human societies and cultures. All information was organized first by culture and then by subject. The innovative feature at the time was that since every paragraph of every piece of descriptive material was subject-indexed, all relevant paragraphs pertaining to the same subject could be "filed" together. HRAF grew out of these earlier efforts. With our online application, researchers can now search by word as well as by subject category at the paragraph-level. HRAF now has two online collections: eHRAF World Cultures and eHRAF Archaeology. HRAF is gradually transforming its paper and microfiche collections to online format. In approximately 5 years, all of the anthropological cultures in the paper and microfiche [HRAF Collection of Ethnography](#) will be online in eHRAF World Cultures.

**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 Universität 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.

**Old World Trade Routes (OWTRAD)**

*URL:* www.ciolek.com/OWTRAD/DATA/oddda.html  
*Sponsors:* Research School of Pacific and Asian Studies, Australian National University, Canberra  
*Principals:* T. Matthew Ciolek  
*Period:* Ancient.  
*Area:* Global.  
*Availability:* Free for individual use.  
*Status:* Ongoing.  
*Description:* Georeferenced historical transport/travel/communication routes and nodes in the Old World.

**Open Street Map**

*URL:* http://en.wikipedia.org/wiki/OpenStreetMap  
*Sponsors:* Wikipedia.  
*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.

**Pleiades**

*URL:* http://pleiades.stoa.org  
*Locations:* the Ancient World Mapping Center at the University of North Carolina at Chapel Hill; the Institute for the Study of the Ancient World at New York University; the Stoa Consortium for Electronic Publication in the Humanities at the University of Kentucky.
Principals: Tom Elliott, Richard Talbert, Roger Bagnall, Sean Gillies.
Period: The Classical age.
Area: The Mediterranean eastward to the Indian subcontinent.
Availability: Free and open-source.
Status: In progress.
Description: Pleiades gives scholars, students and enthusiasts worldwide the ability to use, create and share historical geographic information about the Greek and Roman World in digital form. The project is producing a durable, open-access digital gazetteer for the Greek and Roman world, together with reusable, open-source software necessary to develop, disseminate and collaboratively maintain similar digital publications. Pleiades (named after the daughters of Atlas in Greek mythology) builds upon the NEH and Mellon supported Classical Atlas Project (CAP; 1988-2000), which produced the prize-winning Barrington Atlas of the Greek and Roman World (R. Talbert, ed., Princeton, 2000) and to which the Ancient World Mapping Center (AWMC) at the University of North Carolina in Chapel Hill is heir. With funding from an NEH Preservation and Access Research and Development Grant (January 2006 – June 2008), the AWMC launched Pleiades by prototyping a family of standards-based web services around historical spatial information digitized from a subset of the CAP compilation materials. This information – corresponding to two of the 99 maps appearing in the atlas – includes the locations and attested names of historical and archaeological sites, transportation networks, regions and peoples, together with bibliographic citations leading to additional information about them. In February 2008, the Institute for the Study of the Ancient World at New York University joined forces with AWMC to advance the Pleiades effort; it now serves as the focus of technical development for the project. The Pleiades web services (including a content management system that permits scholars, students and enthusiasts worldwide to suggest corrections and updates to the gazetteer and bibliography) were the major deliverables of the first grant. Additional funding from a NEH/JISC Joint Transatlantic Digitization Collaboration Grant (2008-2009) is expanding the retrospective digitization of CAP content and prototyping standards-based cross-project interoperability mechanisms to support machine-actionable historical citation and geospatial search and visualization functions.

Representing Ethnic Groups in Space
URL: www.icr.ethz.ch/research/greg
Location: ETH-Zurich
Sponsor(s): Nils B. Weidmann, Jan Ketil Rod, Lars-Erik Cederman.
Principal(s): Nils B. Weidmann, Jan Ketil Rod, Lars-Erik Cederman.
Period: Contemporary.
Area: Global.
Availability: Free
Status: Completed.
Description: A geo-referencing of ethnic groups.

Timemap
URL: www.timemap.net
Location: Archeological Computing Laboratory, University of Sydney
Sponsor(s):
Principal(s): Ian Johnson
Period:
Area:
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 Digital Library (WDL)
URL: www.wdl.org/en/
Location: UNESCO.
Sponsors: United Nations Education, Scientific, and Cultural Organization (UNESCO); Google, Inc. ($3 million for the initial development of a WDL plan and the WDL prototype); The Qatar Foundation ($3 million in general support and for the development of the Central Library of the Qatar Foundation as a key node in the WDL network); The Carnegie Corporation of New York ($2 million to support the inclusion of cultural institutions from sub-Saharan Africa and Eurasia in the WDL); The King Abdullah University of Science and Technology, Saudi Arabia ($1 million to support activities relating to the dissemination of digital versions of manuscripts and other
materials relating to science in the Arab and Islamic worlds); Microsoft, Inc. ($1 million in general support); The Lawrence and Mary Anne Tucker Foundation (for the establishment of a digital conversion center at the Iraqi National Library and Archives); The Bridging Nations Foundation (for the development of Middle East-related content).

**Principals:** WDL Working Groups established after the December 2006 Experts Meeting include the Content Selection Working Group and a Technical Architecture Working Group. These groups are comprised primarily of representatives from partner institutions. The International Federation of Library Associations and Institutions (IFLA) and the Library of Congress have co-sponsored a working group to develop guidelines for digital libraries, including the WDL.

**Period:** Unbounded.

**Area:** Global.

**Availability:** Free.

**Status:** Ongoing.

**Description:** [From the web site...] U.S. Librarian of Congress James H. Billington proposed the establishment of the WDL in a speech to the U.S. National Commission for UNESCO in June 2005. The basic idea was to create an Internet-based, easily-accessible collection of the world's cultural riches that would tell the stories and highlight the achievements of all countries and cultures, thereby promoting cross-cultural awareness and understanding. UNESCO welcomed the idea as a contribution toward fulfilling UNESCO's strategic objectives, which include promoting knowledge societies, building capacity in developing countries, and promoting cultural diversity on the Web. UNESCO Director-General Koichiro Matsuura designated UNESCO's Directorate for Communication and Information, led by Dr. Abdul Waheed Khan, to work with the Library of Congress to develop the project. In December 2006, UNESCO and the Library of Congress convened an Experts Meeting to discuss the project. The assembled experts from all parts of the world identified a number of challenges that the project would need to overcome to be successful. They noted that little cultural content was being digitized in many countries and that developing countries in particular lacked the capacity to digitize and display their cultural treasures. Existing Web sites often had poorly developed search and display functions. Multilingual access was not well developed. Many Web sites maintained by cultural institutions were difficult to use and, in many cases, failed to appeal to users, particularly young users. The Experts Meeting led to the establishment of working groups to develop guidelines for the project, and to a decision by the Library of Congress, UNESCO, and five partner institutions - the Bibliotheca Alexandrina, the National Library of Brazil, the National Library and Archives of Egypt, the National Library of Russia, and the Russian State Library - to develop and contribute content to a WDL prototype to be presented at the UNESCO General Conference in 2007. Input into the design of the prototype was solicited through a consultative process that involved UNESCO, the International Federation of Library Associations and Institutions (IFLA), and individuals and institutions in more than forty countries. The successful unveiling of the prototype was followed by a decision by several libraries to develop a public, freely-accessible version of the WDL, for launch at UNESCO in April 2009. More than two dozen institutions contributed content to the launch version of the site. The WDL makes it possible to discover, study, and enjoy cultural treasures from around the world on one site, in a variety of ways. These cultural treasures include, but are not limited to, manuscripts, maps, rare books, musical scores, recordings, films, prints, photographs, and architectural drawings. The WDL Content Selection Working Group initially developed broad guidelines for selection. In addition, WDL partners worked to include important and culturally significant content about every UNESCO member country. The content is in a variety of formats and languages, from different places and time periods. The WDL focuses on significant primary materials, including manuscripts,
maps, rare books, recordings, films, prints, photographs, architectural drawings, and other types of primary sources. One of the WDL's content objectives will be to work closely with UNESCO's Memory of the World program to make publicly accessible digital versions of these collections. Items on the WDL may easily be browsed by place, time, topic, type of item, and contributing institution, or can be located by an open-ended search, in several languages. Special features include interactive geographic clusters, a timeline, advanced image-viewing and interpretive capabilities. Item-level descriptions and interviews with curators about featured items provide additional information. Navigation tools and content descriptions are provided in Arabic, Chinese, English, French, Portuguese, Russian, and Spanish. Many more languages are represented in the actual books, manuscripts, maps, photographs, and other primary materials, which are provided in their original languages. The WDL was developed by a team at the U.S. Library of Congress, with contributions by partner institutions in many countries. The principal objectives of the WDL are: Promote international and intercultural understanding; Expand the volume and variety of cultural content on the Internet; Provide resources for educators, scholars, and general audiences; Build capacity in partner institutions to narrow the digital divide within and between countries.

**World-Historical Dataverse**

**URL:** www.dataverse.pitt.edu  
**Locations:** University of Pittsburgh  
**Sponsor:** University of Pittsburgh  
**Principals:** Patrick Manning, Siddharth Chandra  
**Period:** 1500-present.  
**Area:** World.  
**Availability:** Free.  
**Status:** Ongoing.  
**Description:** In its initial, early stage, the World-Historical Dataverse includes (1) a Gateway to online historical datasets, listing major portals and individual datasets; (2) an Archive for individual datasets; and (3) a Discussion Center including position papers and working papers. Current work emphasizes expanding the archive; modifying archived datasets to include spatial variables in GIS terms; and programming to “federate” individual datasets so that online users can assemble fields from various sources to create new and broader datasets. The project has applied to the National Endowment for the Humanities for support of archiving and federating datasets. Ultimately, the Dataverse project is intended to create a consistent, global, historical database including social-scientific, health, and ecological variables. The Dataverse project is intended to collaborate closely with the Global Historical GIS.

**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

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

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: http://www.flwi.ugent.be/hisgis/
Location: Universiteit Gent, Department of Modern and Contemporary History.
Sponsor(s):
Principal(s):
Area: Belgium.
Status: Ongoing (1990-).
Availability: Available through the internet.
Description: Population, territorial units and their boundaries, as well as agricultural, landuse, and industry statistics are available. Population data are available at the village, municipality, canton, and province levels. Landuse, agricultural, and industrial data are available at the provincial level. The project is described in detail in M. De Moor and T. Wiedemann, “The Historical Geographical Information System of Belgium,” History and Computing 3:1 (2001): 69-96.

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.

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 (HGIS Germany)
URL: www.hgis-germany.de; www.eckompendium-hgis.de
Location: Institute of European History and University of Applied Sciences, Mainz, Germany
Sponsors: Krupp Foundation, State of Rhineland-Palatinate
Principals: Andreas Kunz (Mainz), Alexander Zipf (Bonn/Mainz)
Period: 1820-1914
Area: Germany and central Europe, portions of northern, southern and western Europe
Availability: To be determined.
Status: Completed
Description: HGIS Germany is a multi-purpose historical information system on the forty component sovereign – later federated - states of Germany during the 19th century. Its GIS database contains all boundary changes in Germany from 1820 to 1914 from the state to the governmental district levels on a yearly basis. It is linked to a multimedia database on 250 spatial units, and contains statistical information on population, transport infrastructures, and industrial production.

Historical GIS (HGIS) Research Network
URL: http://www.hgis.org.uk/
**Locations:** Lancaster University (part of the initiative in Digital Humanities).

**Sponsor:** Economic and Social Research Council (ESRC).

**Principals:** Ian Gregory, Paul Ell.

**Period:** Nonspecific.

**Area:** Primarily Europe.

**Availability:** Free.

**Status:** Ongoing.

**Description:** [From the web site...] The Network has four main aims: 1. To provide a focus for Historical GIS research in Europe. 2. To advance our understanding of Historical GIS at the technical, methodological and applied levels. 3. To encourage the adoption of GIS amongst a broad audience of researchers who have an interest in the past. 4. To investigate the setting up of an international association to act as a focus for historical GIS research. The website describes the events organised by the Network, and provides a portal for a wide variety of links and resources relevant to Historical GIS Research including details of training events, a large bibliography, and lists of projects, websites, organisations and software relevant to historical GIS and historical maps.

**Historical GIS for Ireland**

**URL:** http://www.qub.ac.uk/cdda

**Location:** Centre for Data Digitisation and Analysis (CDDA) at the Queen’s University, Belfast.

**Sponsor(s):**

**Principal(s):**

**Period:** 1821-1971

**Area:** Ireland

**Status:** Ongoing.

**Availability:** Data are available through the internet.

**Description:** Data from the Irish Censuses between 1821 and 1971 have been digitized. Administrative boundaries have also been digitized.

**Historical GIS Initiative in Russia**

**URL:** http://www.ihst.ru/personal/imerz/bound/bounds.htm

**Location:** Institute of Geography, Russian Academy of Sciences, Moscow.

**Sponsor(s):**

**Principal(s):** Irana A. Merzlyakova

**Period:** 1775-1920.

**Area:** Russia

**Status:** Ongoing. Currently in a pilot phase.

**Availability:** Data will be made available through the internet.

**Description:** Goal is to digitize the network of Russian administrative boundaries and place names from 1775 to 1920.

**Institute of European History Map Server (IEG-Maps)**

**URL:** www.ieg-maps.uni-mainz.de

**Location:** Institute of European History and Johannes Gutenberg-University, Mainz, Germany.

**Sponsor(s):** Gerda Henkel Foundation.

**Principal:** Andreas Kunz (Mainz).

**Period:** 1350 to the present.

**Area:** Europe, with emphasis on Central Europe and Germany.
Availability: Free.
Status: Ongoing.
Description: IEG-Maps contains over 700 digital maps on the history of Germany, central Europe, and Europe as a whole from c. 1350 to the present. All maps have been newly constructed, no scans are used. Map files can be downloaded and used free of charge for personal, non-commercial use.

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: http://www2.fcsh.unl.pt/atlas/
Location:
Sponsor(s):
Principal(s): Professor Doutor Luis Nuno Espinha da Silveira
Period: 1770-1911
Area: Portugal
Availability:
Status:
Description:

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 Umea 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 Facetted 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 facetted 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.

Towards a Historical Geographic Information System for the Netherlands (HGIN)
URL: http://www.niwi.knaw.nl/nl/geschiedenis/projecten/toon
Location: History Department of the Netherlands Institute for Scientific Information Services (NIWI) in Amsterdam
Sponsor(s):
Principal(s):
Period: 1795-1971
Area: Netherlands
Status: Ongoing.
Availability: The digitized data and maps will be made freely available on the internet.
Description: The project is in the process of digitizing all Dutch historical censuses, and the accompanying maps.

EAST ASIA

China in Time and Space (CITAS)
URL: http://citas.csde.washington.edu/
Location: University of Washington.
Sponsors: [see below]
Principals: A partnership of China scholars, sponsored by the Joint Committee on Chinese Studies (JCCS), and the Consortium for International Earth Science Information Network (CIESIN). CITAS database management and technical work takes place at several sites, including the University of Washington, the University of Michigan, the University of California at Davis, and at CIESIN in Saginaw, Michigan.
Period: Ancient times to present day.
Area: China.
Availability:
Status: Ongoing.

Description: [From the web site...] China in Time and Space (CITAS) is a project to create and maintain databases of spatially- and temporally-referenced data on China. CITAS aims to provide access to a wealth of contemporary and historical information on China at minimal cost to scholars and other non-commercial users. The project utilizes state-of-the-art geographic information system (GIS) technology to integrate tabular and vectorized map data. An unusual aspect of CITAS is its emphasis on the temporal dimension: one project goal is the capability of reconstructing Chinese administrative geography for specific historical dates, and the ability to compare spatial data at two points in time, whether over years or over centuries. Collaboration with Chinese scholars and organizations is integral to the project goals.

Chinese Civilization in Time and Space
Location: Academia Sinica, Taiwan.
Sponsor: Academia Sinica, Taiwan.
Principals: Fan I-chun, Associate Research Fellow, Institute of History and Philology, Academia Sinica; Simon C. Lin, Director, Computing Center, Academia Sinica; Eric Yen, Maintenance Engineer, Computing Center, Academia Sinica.
Period: Two millennia.
Area: China.
Availability: Restricted. [From the web site...] The copyright of the Tan's Atlas belongs to the Chinese Academy of Social Sciences. The Academia Sinica has been exclusively authorized to vectorize it, and no other person or organization is allowed to do so again. The copyright of the vectorized Tan's Atlas belongs to the Academia Sinica and the Chinese Academy of Social Sciences. The copyright of the Arc/China belongs to the National Bureau of Surveying and Mapping, P.R.C. The Academia Sinica has been authorized to integrate it with the vectorized Tan's Atlas, and makes it available to the specific organizations and persons for academic purpose. The copyright of the above mentioned dataset should not be infringed in any form. Users of the CCTS should acknowledge the copyright as the follows: Academia Sinica, Chinese Civilization in Time and Space, Version 1, (Taipei, September 2002).
Status: Ongoing.

Description: [From the web site...] The creation of the Chinese Civilization in Time and Space (CCTS) was mainly based on Professor Tan Qixiang's The Historical Atlas of China (Tan's Atlas) and the 1:1,000,000 Digital Map Database of China (Arc/China). In response of the demands of interdisciplinary research applications, this system aims to construct an integrated GIS-based application infrastructure within the spatial extent of China, in the timeframe of Chinese history, and with the contents of Chinese civilization. Although target users are primarily scholars, academic experts, and school teachers, it is also our goal to provide general spatial- and temporal-based applications. This, we hope, will establish a feedback mechanism of continuously collecting research and application results to fertilize the contents of this system and elaborate the value of information integration. This system consists of three major components: basic geospatial materials, WebGIS integrated application environment, and thematic information. The fundamental base maps are based on Dr. Tan's "The Historical Atlas of China". "The Historical Atlas of China" provides users with Chinese historical features, covering Chinese history over the past 2000 years, from the ancient time to Qing dynasty. Furthermore, various historical atlas and remote sensing imagery are persistently geo-referenced and overlaid into the system to broaden the
spatial and temporal scope. System features (1) A WebGIS-based distributed application infrastructure is constructed to provide both content navigation and Web mapping. (2) Lower the cost for doing things in GIS way, and with shorter learning curve. (3) Capable of integrating any geospatial and attribute information in the Internet. (4) Users could upload their results and contribute to the system, or just save to their local clients alternatively. (5) Having scalability, integration and security consideration into the design. (6) The systems of Scripta Sinica (漢籍電子文獻系統) has been integrated with CCTS. A virtual Sinological research environment could be fulfilled by continuous integration of all kinds of public information system and research outcome.

A Searchable Relational Database of Buddhist Monasteries in China (BGIS)

URL: http://bgis.coh.arizona.edu/
Location: University of Arizona.
Sponsor: Vice President for Research and Graduate Studies, the University of Arizona.
Principal: Jiang Wu, Assistant Professor, East Asian Studies, University of Arizona.
Period: 17th century (initially).
Area: China.
Availability: Public, when complete.
Status: Ongoing.
Description: [From the web site...] The proposed project aims to build a searchable relational database of Buddhist monasteries in China by using Geographic Information System (GIS) tools. Combining the power of databasing and electronic mapping, this software can integrate different types of data about monasteries and illustrate them visually on maps. This database will allow researchers to do quantitative studies about the geographical distribution of Chinese Buddhism, the formation of pilgrimage networks, and their relationship to other Chinese cultural and social institutions. This project has developed from my research on seventeenth-century Chinese Buddhism, which focused on the rise of a new sect in Fujian and Zhejiang provinces. The regional development of this sect brought my attention to the geographical distribution of religious institutions and their relation to other cultural and social institutions. Although a large amount of data about these monasteries exists, no one has attempted to decipher the relationship among sacred geography, natural landscape, pilgrimage networks, and political and social institutions. GIS is an efficient tool to reveal the geographic pattern of religious distribution and it has the potential to add a new dimension to the way we study religion. Traditionally scholars of religion have relied on textual scholarship without the benefit of quantitative and computerized research methods. But through GIS, empirical patterns and spatial relationships hidden in massive amounts of data become evident and visualized. This project will not only provide a database of Buddhist institutions in China, but will also apply a new methodology to religious studies. In my earlier work, I have begun to collect three types of data about Buddhist monasteries: genealogical records of abbots, biographies of clergy, and local gazetteers that record detailed information about the succession of abbots and their sectarian affiliations. These data enable me to map three key elements of the monastic network in China by using GIS: the location of monasteries, the abbots of monasteries in a given year, and their sectarian affiliations. In association with other geographical and cultural parameters such as main transportation routes, merchant networks, and the number of exam degree holders, the powerful analytic tools provided by ArcView, the most sophisticated software developed by Environmental Systems Research Institute (ESRI), will allow me to analyze the relationship between the geographical distribution of the monastic sites, the pilgrimage network, and regional cultural and economic development. The first phase of this
project is to design a pilot database with the help of a student worker who is specialized in GIS modeling and our consultant Lex Berman, who is project manager of Harvard University China GIS. Their job is to devise a workable database model according to the relationships of different types of data and to deal with potential technical problems. In the meantime, to complete the data concerning monasteries, I plan a trip to the Library of Congress that holds the most complete collection of local gazetteers of Chinese monasteries. (These sources are rare books and are not accessible through interlibrary loans.) Then Chinese-English bilingual student workers will input these data by using Microsoft Access. The software for inputting Chinese character is NJStar which allows Unicode input. (Due to the large amount of data, this project will first concentrate on data concerning monasteries in seventeenth-century China.) The primary task is to input three types of information: location of monastery, name of the presiding abbots, and their sectarian affiliation. In addition, this GIS database will also document architectural and natural features with digital photography if these data are available. Furthermore, for famous monasteries, portions of local gazetteers relevant to the monasteries and notes about available studies on these monasteries will be incorporated. These data will be available and searchable in translated English, Pinyin romanization, and classical Chinese characters. After the completion of data input in Microsoft Access, I will connect the database with ArcView and ArcInfo. After testing and revising, this GIS system should be searchable by year, location, monastery, and abbot name in both English and Chinese. It will be published on a project website designed by student workers and produced in 10 additional copies to be distributed to other interested scholars and institutions for further testing and revision.

**Japanese Historical Maps**

**URL:** http://www.davidrumsey.com/japan/

**Location:** East Asian Library, University of California at Berkeley.

**Sponsor:** Cartography Associates and the East Asian Library.

**Principals:** Peter Zhou, Director of the East Asian Library; David Rumsey.

**Period:** 17th-20th centuries.

**Area:** Japan.

**Availability:** On-line.

**Status:** Ongoing.

**Description:** Maps are geo-referenced but no data is extracted. The Japanese Historical Map Collection contains about 2,300 early maps of Japan and the World. The collection was acquired by the University of California from the Mitsui family in 1949, and is housed on the Berkeley campus in the East Asian Library. Represented in this online collection are over 1100 images of maps and books from this Collection. The maps were selected by Yuki Ishimatsu, Head of Japanese Collections at the East Asian Library, and scanned and put online by David Rumsey and Cartography Associates. The project was initiated by Peter Zhou, Director of the East Asian Library. Funding and project management is provided by Cartography Associates and the East Asian Library. When the University of California at Berkeley purchased the Mitsui Library from the Mitsui family in 1949, included among the 100,000 items was a collection of 2,298 maps which had been assembled by Mitsui Takakata (penname: Soken) (1882-1950), the 9th head of the Shinmachi branch of the family. The most unusual part of the collection is the 697 woodblock-print maps (and a few dozen manuscript maps) dating from the Tokugawa period (1600-1867). Especially rare is the collection of Tokugawa city maps: 252 maps of the city of Edo (modern Tokyo), 79 maps of Kyoto, 40 of Osaka, and 30 maps of other cities such as Kanazawa, Nagoya, Nagasaki, and Yokohama, all from the Tokugawa period. Among the earliest maps are those of Osaka (1656),
Kyoto (1654-68), and Edo (1676). The earliest Japanese world maps also date from this period. In collecting maps from the Meiji period (1868-1912), Mitsui Soken also displayed his antiquarian interest by concentrating on the earlier decades. Most of the Meiji maps date from the period before 1890 and many are printed on handmade paper; a considerable number are printed from woodblocks. Meiji period city maps include 240 of Tokyo, 112 of Kyoto, 89 of Osaka, and 312 of other cities. A significant number of the maps in the collection, and many of the world maps, are in color.

**HGIS Project of Korean Culture**

*URL:* http://ikc.korea.ac.kr/culture/culture_start.htm  
*Location:* Institute of Korean Culture, University of Korea.  
*Sponsor(s):*  
*Principal(s):* Jong-Hyuk Kim.  
*Period:* 1392-1910 (Chosun Dynasty).  
*Area:* Chosun Dynasty (South Korea).  
*Status:* Ongoing.  
*Availability:*  
*Description:* Project began in August 2002. Research team consists of thirty professors and thirty graduate students from the fields of geography, history, philosophy, literature and folklore. The project’s goal is to digitize cultural information which will then be transferred into maps, creating digital ‘culture maps’.

**NORTH AMERICA**

**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 (see Japanese Historical Maps, listed elsewhere).  
*Availability:* A small portion of the 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.

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.

HGIS Print Culture of Canada
URL:
Location: Dalhousie University
Sponsor(s):
Principal(s): Fiona A. Black and Bertrum H. MacDonald
Period:
Area: Canada
Status: Ongoing.
Availability: Currently at a limited access site, but is planned to be made publicly available.
Description: Collecting historical GIS information related to historical data on print culture within Canada.

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 D:
Historical Atlases

Canada

United States

North America

Central America

South America

Latin America

Central/Eastern Europe


**Europe/Near East/Middle East**


Hertslet, Edward. 1875-1891. *The Map of Europe by Treaty; Showing the various political and territorial changes which have taken place since the general peace of 1814*, 4 vols. [Available on Googlebooks. Scans of parts of it available at: http://www.maproom.org/00/02/index.php.]


*Tabula imperii Byzantini*, Vienna 1976ff (published in Denkschriften, Österreichische Akademie der Wissenschaften, phil.-hist. Klasse, 125, 131 etc.).


**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**

Anderson, Charles S. *Augsburg Historical Atlas of Christianity in the Middle Ages and Reformation.*


**Islam, Muslims**


**Language**


Appendix E:
Non-Cartographic Data Sources

General


*The Central Eurasian Information Resource*. A uniquely cataloged collection of information resources, some developed at the University of Washington (examples are described below), others acquired contractually (such as the UW Libraries' subscriptions to online Russian and East European newspapers), and the remainder freely available on the Internet, but identified as particularly valuable by subject and area specialists, and selected for inclusion in the CEIR Online Resource Collection on the strength of their quality. The collection is expected to number between 500 and 1,000 cataloged resources by the completion of the current stage of the CEIR project in late 2002, and it will be added to continually thereafter. The CEIR Resource Collection is cataloged using a set of controlled metadata that differs from conventional library metadata, and it is searchable through a specially designed portal. [http://green.lib.washington.edu/website/ceir/]


63
Europa Yearbook, 1926-.
“European State Finance Database.” [Comprises dozens of datasets housed at the Arts and Humanities Data Service (AHDS), including revenue and expenditure data for countries and cities, mostly though not exclusively in Europe, back to the Middle Ages.]
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.


[Includes bilateralism data, commodity trade data, exchange rates (annual), exchange rates (weekly) and tariff indicators, from 1900-1939.]


Lindert, Peter. Data on life expectancy and infant mortality. Stored at GPIH:
http://gpih.ucdavis.edu/Evidence.htm

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.]


http://www.econ.brown.edu/fac/Louis_Putterman/

*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 Office Lists,* London: Harrison, 1878-1940.

*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.*

**Chinese Empire/China**

*The Online Index of Chinese Buddhism.* An annotated list of online resources for the study of Buddhism in China [http://www.columbia.edu/~gas2122/oicb.html]

**French Empire/ France**

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

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


German Empire/Germany
Deutscher Kolonial-Kalender

Ottoman Empire/Turkey
Cuinet, Vital. 1892. La Turquie d’Asie: Géographie Administrative, Statistique. E. Leroux.

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
Repartição de Estatística Geral: Anuario 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: Anuario Estatístico 1933-1952

Central African Republic
Direction de la Statistique et de la Conjoncture: Annuaire Statistique de la République Centrafricaine 1962

Chad

Congo (formerly French Congo)

Egypt (See also United Arab Republic)
Maslahat al-Ihsa wa-al-Ta'dad: 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)

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 Mécanographie: *Annuaire Statistique de la République du Mali 1963-1973*

Mauritania
Direction de la Statistique et des Etudes Economiques: *Annuaire Statistique 1968-1974*

Mauritius
Central Statistics Office: *Ar Book of Statistics 1946-1959*
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
Repartiçao Técnica de Estatística: *Annuario Estatistico/ 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*
Western State. Statistics Division: *Statistical Abstract 1959-1972*

Senegal
Direction de la Statistique: *Situation Economique du Sénégal 1962-1976*

Sierra Leone
Central Statistics Office: *Annual Statistical Digest 1968-1976*

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

South Africa
Office of Census and Statistics: *Union Statistics for Fifty Years: Jubilee Issue, 1910-1960*
Department of Statistics: *Statistical Year Book 1964-1966*
*South African Statistics 1968-1976 (Sheehy CB210)*

Spanish Sahara
Secretário General: Sahara Espanol: *Annuario Estadistico 1948-1950*

Swaziland

Tanzania (formerly Tanganyika and Zanzibar)

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
Statistics Division: *Statistical Abstract 1957-1973*

United Arab Republic
Central Agency for Public Mobilisation and Statistics: *Statistical Handbook of the United Arab
Republic 1952-1971
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