

On the Cyberinfrastructure for GIS-Enabled Historiography
Peter K. Bol
Harvard University, Center for Geographic Analysis

Abstract:

From an historian's perspective the use of GIScience and technology in the study of history holds the promise of an integration of historical and geographic modes of analysis. The national geographic information systems that provide extensive coverage of changes in administrative structures over time provide important support for GIS-enabled historiography. Other parts of the cyberinfrastructure necessary to support collaborative research in a digital environment are now beginning to emerge, however a world-historical gazetteer, an essential tool for linking historical data to mapped places has yet to be developed.

History as a record of the past tracks change; history as a discipline turns to new ways of understanding the past. There was once a quantitative turn, a cultural turn, a linguistic turn, and today there is a “spatial” turn. “Spatial turn”—the recognition that knowing where things took place is necessary to understanding what took place—has not yet become as widely used as “cultural turn” but recently it has been outpacing “spatial history” (according to a Google Ngram search). In fact the spatial turn was announced over a decade ago with Anne Knowles’ special issue of Social Science History on historical geographic information systems (GIS) (Knowles 2000), followed by successive collections illustrating the application of GIS to historical subjects (Knowles 2002; Knowles and Hillier 2008). Space, place, and landscape have regularly figured in modern thinking about the past (Guldi 2011), although not, it seems to me, with the systematic analysis of the spatial dimensions of change over time that is integral to spatial history (White 2010).

The kind of spatial history that makes use of GIS is one particular way of making the spatial turn, a turn that some argue has been taking place across the disciplines (Guldi 2011; Warf and Arias 2009). To be clear, this is but one possible marriage between history and geography. The spatial turn also appears in studies focused on “place,” as sites where social processes and events take place and places that serve as the basis for identity which constructed through social processes (Withers 2009). A reliance on GIS differentiates this kind of historical study from humanistic geography or the geographically engaged research of the “geo-humanities” (Dear et al. 2011). What is clear is that during the last decade we have begun to see major historical studies where the research findings depended on geospatial analysis (Ayers and Rubin 2000; Ayers 2003; Gordon 2008; White 2011).

An awareness that change over time happens in many places across space, and that spatial features are not constant through time is certainly ancient, but geography and history for all their overlaps have remained distinct. The polymaths Strabo and Sima Qian, to give ancient examples from opposite ends of Eurasian landmass, were near contemporaries (1st century BC). Strabo is known to us as a geographer (but he was interested in philosophy and astronomy as well and also wrote a work known as the

“Historical Sketches”). Sima Qian is treated as an historian (but he was also interested in philosophy, astronomy, and geography). They both recognized that life unfolds over time and in many places at once; from a geohistorical perspective it is temporal and spatial. How did they model this (I use model because that is what Sima and Strabo were in fact doing)? Both turned away from the possibility of a quantitative, “scientific” methodology, which was proposed at the time in both places by some scholars who looked to mathematics and astronomy for methods to model space and time. Ultimately they went in different directions. Strabo devoted years to his *Geography*, in which he described the world known to him through its diverse parts (he included South but not East Asia) and Sima Qian, who gave years to his *Records of the Historian*, documented the diversity of the past through court-centric chronologies and extensive biographies of individuals from across the land.

The real and difficult challenges of combining time and space, of the map and the chronology, are themselves the subject of research (Peuquet 1994; Peuquet 2002; Andrienko et al. 2010). Yet historians and geographers face a similar challenge: they are always engaged in choices of scale. In theory time is infinite and geospace is finite (or historical time is finite and space is infinite), but in practice human historical time and geographical space are finite yet subject to resolution at ever-finer scale. Historians and geographers are respectively engaged in representing temporal change and spatial variation, and in simplification, in order to establish analytic clarity. A chronology, integral to history, and a map, integral to geography, are useless if perfectly to scale. We choose to highlight what is important: to widen the road so that it is visible, to define certain moments as consequential in order to clarify change and difference.

History and geography have been, and remain, separate disciplines, evident in the subtitle of Alan Baker’s *Geography and History: Bridging the Divide* (Baker 2003). We can ask why this should be so – which is to ask why methodologies for analyzing change over time and variation through space have diverged so greatly – but I prefer to ask, in the spirit of the sixty sessions on “Space-Time Integration in Geography and GIScience” at the 2011 AAG meeting, what the technologies of today allow us to do about it. The great modern advancement of knowledge has been credited to three things: academic specialization, paradigm shifts, and the emergence of new tools. For the moment I am going to stand with the “tool” camp, and suppose that tools that enable us to deal with vast quantities of information allows us to see with many times (for historians) and many places at once (for geographers) affect both specialization and paradigm shifts. GIS as a tool, like the telescope and the microscope, allows us to see what we could not see before. Historical GIS may not bring about the integration of history and geography but it does make it possible for historians to take advantage of some of the accomplishments of GIScience to combine variation through space with change over time.

Learning to apply GIS software to historical questions may not be too demanding, thanks to GIS textbooks and Ian Gregory’s guide to historical GIS specifically (Gregory and Ell 2007), but building comprehensive historical geographic information systems is very challenging and time-consuming, as the review of national historical GIS projects in the journal *Historical Geography* makes clear (Knowles 2005). GIS gives us the ability to

sort out and analyze the ever increasing amount of digital information with spatial attributes, but without the comprehensive historical geographic information systems that have been and are being constructed it would be impossible for the individual researcher to handle national datasets, visualize and measure the spatial patterns in that data, see those patterns change over time and correlate information from different domains.

The barriers for historians in applying GIS technology to particular questions are not so much technological as infrastructural. The American Council of Learned Societies (ACLS) report on cyberinfrastructure for the humanities and social sciences asserted that it was “primarily concerned not with the technological innovations that now suffuse academia, but rather with institutional innovations that will allow digital scholarship to be cumulative, collaborative, and synergistic.” (Welshons and al. 2006). How does this apply to using GIS in service of history?

We can begin with one of the first steps in the research cycle: finding information. GIS is about geographic space, but in the historical record “place” not “space” was the focus. Populations are clustered in places, people come from places, and postal stations are places in themselves. Places are nodes in networks, but the further back we go the less certain our knowledge of the precise routes between nodes than our knowledge of where the nodes/places were. Reliable sources for boundaries only begin to emerge with the spread of modern cartography in the nineteenth century.

Finding information has more to do with place than with space, very much in the sense of Curry’s distinction (Curry 2005). Historical information, particularly for periods lacking cartographic records, is typically associated with places—a person of a certain place, a religious site located at a certain distance from a known place, the tax assessment of a place—although a spatial picture can often be inferred. The interface humans have created between themselves and the physical world, which allows them to position themselves relative to that geography and make it intelligible, to organize knowledge of it, and preserve the memory of their acts within it, is created through the process of naming. Naming – of a mountain or a river, town or a building – maintains an intelligible interface between the geophysical world and human culture; the name makes it a place. Naming pertains to all aspects of human life. But, like everything in human culture, names are not stable. They are changed, abandoned, forgotten, fabricated.

Names provide an interface between the historical and the geographic, allowing us to use them to bring the historical and geographic together. If we can capture names in written/drawn sources and locate those on the landscape, then we can locate historical data (tax records, population, religious activities, battles) in space. In this ideal version, in which all temporal data can be linked to places and places to spatial objects, the data from past becomes attributes space and place. The historical record is, ultimately, finite, so it is possible to imagine collecting all names in one giant historical gazetteer that tells us when a name is valid, what system of naming it belongs in, and (we hope) where it is. The gazetteer is fundamental to the geographic ordering of our human past and making it accessible. In practice, however, a gazetteer should be able to accommodate place names (even imaginary and fabulous landscapes) that have sources but lack spatial locations.

The United States government is the most important source of geographic names in the world today. The Geographic Names Information System (GNIS) provides over 2 million names for named natural and constructed places (except roads) in the United States (Yost and Carswell 2009). The National Geospatial-Intelligence Agency's GEOnet Names Server (GNS), the official repository for foreign place names, has over 5 million foreign geographic features, including alternate names and the local vernacular. (National Geospatial-Intelligence Agency 2012). These invaluable resources, freely available for public use, are under the aegis of the Board of Geographic Names which has ultimate authority over the names included. The greatest non-governmental gazetteer, the GeoNames geographical database, founded by Mark Wick, contains over 10 million geographical names and consists of 7.5 million unique features whereof 2.8 million are populated places and 5.5 million are alternate names (Wick 2012). A vital difference with the government gazetteers, whose data it incorporates, is that GeoNames allows volunteered data and is open to the wisdom of crowds.

However, none of these gazetteer databases includes time as an attribute of place. At first glance this may appear to be a minor loss—the proportion of the 20 million daily web services requests to GeoNames that need dates is probably very small. But the problem of excluding temporal data for names—including contemporary names—has consequences. The lack of a record about when a name is changed or a jurisdictional line redrawn eventually will result in the loss of knowledge about when the attributes of that place name (population, area, etc.) are valid. In the past territorial administrations managed their records through print, archiving past records and thus creating a paper trail; to the degree that an information management system keeps itself up-to-date by overwriting earlier data that information system is sacrificing a longitudinal record to clerical efficiency. Thus a first-order cyber infrastructural need in integrating history and geography, time and space, is a temporally-enabled gazetteer—in short we need a world historical gazetteer.

The challenges here are considerable. Names can be linked to geographic locations but sometimes only to areas (within the jurisdiction of, mentioned as being near to). Place names change asynchronously. The recoverable begin and end dates are often approximate. There is an enormous amount of information to be had from online resources such as Wikipedia in many languages. How do we organize linked data into geohistorical factoids?

A world-historical gazetteer is fundamental to historically-conscious spatial research. As Humphrey Southall has written:

“Understanding the larger socio-economic challenges facing our society requires a long-term global perspective, but in practice such perspectives are almost impossible to achieve because the necessary datasets are fragmentary or non-existent. All too often, historical research is based on a single country or a small group of advanced economies; or on just the last thirty or forty years. We need to assemble not just historical statistics but closely integrated metadata, including locations and reporting unit

boundaries, so that researchers can explore alternative approaches to achieving consistency over space and time without requiring an army of assistants for each new project...existing social science data repositories are insufficiently integrated...an open collaborative approach is essential...Geographical Information Science technologies are necessary...and concepts from other areas of Information Science are also needed, notably including ontologies and linked data.” (Southall et al. 2011)

But what a world-historical gazetteer should contain and how it should be organized is not settled. The panels on gazetteers for the Space-Time Symposium at the 2011 AAG addressed ontologies of place, temporal frameworks for gazetteer elements, the construction of historical and cultural gazetteers, interoperable gazetteers and the spatially enabled web, building world historical gazetteers from historical GIS, data models and content standards, and building a temporally enabled global gazetteer. The goals of these sessions, as the principal organizers Merrick Lex Berman and Humphrey Southall explained, was to evaluate current gazetteers, to consider methods for building temporal/historical gazetteers, and to persuade the agencies responsible for authoritative gazetteer systems to include time as an essential element. The overall aim was to plan for system interoperability between online gazetteers, and to sketch out the right course of development leading to the funding of a true world historical gazetteer system.

A number of national historical GIS databases provide the kinds of information that a true temporally enabled world gazetteer would need to offer. The Great Britain Historical GIS, which covers the last two centuries of administrative units and the relationships between them is also a sophisticated historical gazetteer(2001-; Southall 2003), accessible through the A Vision of Britain through Time website . The Great Britain Historical GIS (GBHGIS) was first created to enable the longitudinal spatial analysis of demographic data, but precisely because it is also a gazetteer it can link to other data sources: historical maps, election results, and travel writing. Similarly the National Historical GIS was created for the spatial analysis of United States census data 1790-2010, but its polygons have temporal attributes and could provide US data for a historical gazetteer together with the invaluable print *Historical Gazetteer of the United States* (Hellmann 2005). Somewhat similar are the historical GIS of Belgium, Netherlands and Germany (DANS and Universiteit 2006; Kunz, Zipf, and Böhler 2008-; Vakgroep nieuwste geschiedenis/Department of Modern History at Ghent University).

In contrast, the China Historical GIS (CHGIS), covering 221 BC- AD1911 was created as a time series of administrative entities and major towns and their changing relationships between places from 221 BCE to 1911CE. CHGIS from the start has served as a gazetteer in that the purpose was to provide the points and polygons for places to which scholars could join historical data of their choice (Bol 2007; Bol and Ge 2005; Bol et al. 2002-2012). The AAG’s Historical GIS Clearing House and Forum provides a listing of historical GIS projects and gazetteers (Association of American Geographers).

The cyberinfrastructural challenge is to create either a unified or a federated temporally-enabled multilingual gazetteer system informed by multiple ontologies in different languages that can be sustained over time.

This leads directly to a second challenge: populating a world historical gazetteer systematically on a large scale. At first glance the problem is so large that it is hard to say where to begin. There are, I think, two somewhat different starting points: digital texts and scanned, georeferenced maps. The identification of place names appearing in dated texts provides a source authority for a “before” date for a place name. The proprietary Metacarta Geographic Search and Referencing Platform from QBase appears to be the most sophisticated geo-referencing software, which presumably could be used for the geo-tagging and their geo-referencing of historical texts (although with greater degrees of uncertainty as distance from the present increases). Nevertheless, identifying all the place names in past writings provides a large amount of raw data; the hope is that their locations can gradually be refined through iterative procedures.

Since the use of theodolites in 1790s Britain, mathematically accurate maps have accumulated and now cover the entire globe. These maps provide information routes, boundaries, physical features, and locations that texts cannot provide. For a limited historical period – but one which saw global modern growth at a pace unparalleled in human history—geo-referenced maps allow us to link place names, locations, and time and thus provide a foundation for geo-referencing place names that appear in earlier texts. Manual data extraction will always be limited to specific projects; a systematic approach requires the extension of optical character recognition technology to maps. This has largely eluded software engineers but real progress is being made (Chiang and Knoblock).

Given software to extract vector and text data from map scans, a third infrastructural challenge follows: creating a system for discovering and accessing geo-referenced map scans. The premier online collection of scanned maps, with over 29,000 out of a total collection of over 150,000 maps, is the Rumsey Historical Map collection (Rumsey 1996-). Of the scanned maps some 22,000 have rough geo-referencing of which 1000 have been georectified using an average of 20-50 control points per map. Some universities have larger map collections (Harvard has over 500,000 items) but none can rival Rumsey for digitized maps and geo-referenced maps. University map collections do not necessarily register their entire holdings in electronic catalogs, making a union catalog impossible. Given the costs of scanning and geo-referencing the maps in public and private collections, there is a need for a federated system for registering maps that have been scanned or geo-referenced.

Both raster and vector data belong in a federated geospatial catalog. Here there is good news to report. Harvard, MIT, and Tufts have joined in OpenGeoportal.org, to create a portal for searching and previewing collections that can be installed on local servers (it has already been adopted by fifteen universities and government organizations). System interoperability between the portals of different collections will enable searching across catalogs.

A concomitant to a federated spatial catalog is a system for archiving and searching historical datasets, exactly what spatial historians could join to GIS boundary and point files. The Center for Historical Information and Analysis, directed by Patrick Manning at the University of Pittsburgh, has launched the World-Historical Dataverse with the aim of creating such a system and founded the electronic *Journal of World-Historical Information* (World-Historical Dataverse).

The final piece of cyberinfrastructure is an online platform for sharing spatialized data. Here too there has been significant progress. Google Earth created a foundation of public understanding and an inspiration for further developments aimed at research and teaching. Social Explorer, led by Andrew Beveridge, is a proprietary platform with free and subscription editions for the visualization of spatialized data. It includes a wide variety of historical and modern data from the U.S. Census, the American Community Survey, and data on religion. It allows users to create reports, download data in convenient formats, and create a time series of map visualizations (Social Explorer).

ESRI's proprietary freeware, ArcGIS Online and ArcGIS Explorer Online, are cloud-based geospatial content management systems for storing and managing maps, data, and other geospatial information (ESRI). They allow users to create and share maps and datasets, manage geospatial content, and control access to volunteered content. . Another similar system is GeoCommons developed by the GeoIQ company, which has now been bought by ESRI (GeoIQ).

The Center for Geographic Analysis at Harvard is developing the open-source and open-access WorldMap platform to lower barriers for scholars who wish to explore, visualize, edit, collaborate with, and publish geographically referenced information. WorldMap has an expanding list of functionalities it wishes to add, but at this writing it already allows researchers to upload large datasets and overlay them with their own layers or those shared by others, create and edit maps and link map features to rich media content, grant edit permission to small or large groups, export data to standard formats, georeference paper maps scans online, and publish data to a few collaborators or the world (Center for Geographic Analysis).

Harvard's instance of WorldMap runs on Amazon Web Services, although it can be installed on local servers. It is simple to replicate for organizations that would like to set up their own instances cost effectively. Using the Harvard instance, users can upload their data for sharing and archiving through Harvard, and link to external web services. The great promise of WorldMap is that it is cumulative, and this is already being borne out: from its Beta release in July of 2011 to February 2013, it attracted users from than 200 countries, and its over 7000 registered users had contributed 7500 data layers and created 1800 map collections.

The spatial turn in history points toward bringing history and geography together in ways that are changing the ways historians work. Much of it takes place in a digital environment; it involves historians with geographers; it requires collaboration between academics, technologists and librarians; and it must be a cumulative enterprise where we

all advance by sharing our data. The elements of cyberinfrastructure discussed here will help make large-scale historical GIS possible.

- Andrienko, G., N. Andrienko, U. Demsar, D. Dransch, J. Dykes, S.I. Fabrikant, M. Jern, M.J. Kraak, et al. 2010. Space, time and visual analytics. *International Journal of Geographical Information Science* 24: 1577-1600.
- Association of American Geographers 2010-. Historical GIS Clearing House and Forum. http://www.aag.org/cs/projects_and_programs/historical_gis_clearinghouse (last accessed 18 February 2013).
- Ayers, Edward L. 2003. In the presence of mine enemies: war in the heart of America, 1859-1863. The valley of the shadow project. New York: W.W. Norton.
- The Valley of the shadow: two communities in the American Civil War.* Macintosh/Windows version. (1 computer optical disc). W.W. Norton & Co., Charlottesville, Va.
- Baker, Alan R. H. 2003. *Geography and history: bridging the divide. Cambridge studies in historical geography* 36. Cambridge, U.K.; New York: Cambridge University Press.
- Bol, Peter K. 2007. "Creating a GIS for the history of China." In *Placing history: how maps, spatial data, and gis are changing historical scholarship*, eds. Anne Kelly Knowles and Amy Hillier, 25-57. Redlands, CA: ESRI Press.
- Bol, Peter K., Jianxiong Ge, Merrick Lex Berman, and Zhimin Man. 2002-2012. *China Historical Geographic Information System* 1.0-5.0, <http://www.fas.harvard.edu/~chgis/> (last accessed 18 February 2013). Harvard University and Fudan University.
- Bol, Peter K., and Jianxiong Ge. 2005. China historical GIS. *Historical geography* 33: 150-2.
- Center for Geographic Analysis. *WorldMap* and *MapWarper*. Harvard University, Cambridge MA. <http://worldmap.harvard.edu/> and <http://warp.worldmap.harvard.edu> (last accessed 18 February 2013).
- Chiang, Yao-Yi, and Craig A. Knoblock 2009. "Recognition of Multi-Oriented, Multi-Sized, and Curved Text." In *Proceedings of the Tenth International Conference on Document Analysis and Recognition*. <http://www.icdar2009.org/> (last accessed 18 February 2013).
- Curry, Michael R. 2005. Toward a geography of a world without maps: lessons from Ptolemy and postal codes. *Annals of the Association of American Geographers* 95 (3):680-691.
- DANS, and Afdeling Geschiedenis van de Radboud Universiteit 2006. *NLGis*. <http://nlgis.dans.knaw.nl/HGIN/Home.ctrl> (last accessed 18 February 2013).
- Dear, Michael, Jim Ketchum, Sarah Luria, and Doug Richardson. 2011. *Geohumanities: art, history, text at the edge of place*. London and New York: Routledge.
- ESRI. *ArcGIS Online* and *ArcGIS Explorer Online*. Redlands, Calif. <http://www.arcgis.com/home/> (last accessed 18 February 2013).
- GeoIQ. *GeoCommons*. <http://geocommons.com/> (last accessed 18 February 2013).

- Gordon, Colin. 2008. *Mapping decline: St. Louis and the fate of the American city, Politics and culture in modern America*. Philadelphia: University of Pennsylvania Press.
- Great Britain historical geographical information system. *a vision of Britain through time* <http://www.visionofbritain.org.uk/> (last accessed 18 February 2013).
- . 2001-. *Great Britain historical geographical information system*
- Gregory, Ian, and Paul S. Ell. 2007. *Historical GIS : technologies, methodologies and scholarship*. Cambridge studies in historical geography 39. Cambridge, UK; New York: Cambridge University Press.
- Guldi, Jo. The Spatial turn in history. Institute for Enabling Geospatial Scholarship at the Scholars' Lab at the University of Virginia Library, 01 APR 2011 2011. Available from <http://spatial.scholarslab.org/spatial-turn/disciplinary-perspectives/the-spatial-turn-in-history/> (last accessed 18 February 2013).
- Hellmann, Paul T. 2005. *Historical gazetteer of the United States*. New York: Routledge.
- Knowles, Anne Kelly. 2002. *Past time, past place: GIS for history*. Redlands, Calif.: ESRI Press.
- . 2005. Emerging Trends in Historical GIS. *Historical Geography* 33: 7-13.
- Knowles, Anne Kelly 2000. Special issue: historical GIS: the spatial turn in social science history. *Social Science History* 24.3.
- Knowles, Anne Kelly, and Amy Hillier. 2008. *Placing history: how maps, spatial data, and GIS are changing historical scholarship*. Redlands, Calif.: ESRI Press.
- Kunz, Andreas, Alexander Zipf, and Wolfgang Böhler. 2008-. *Historical GIS of Germany* <http://www.hgis-germany.de/> (last accessed 18 February 2013). Institut für Europäische Geschichte Mainz.
- National Geospatial-Intelligence Agency. 2012. GEOnet Names Server <http://geonames.nga.mil/ggmagaz/> (last accessed 18 February 2013).
- Peuquet, Donna J. 2002. *Representations of space and time*. New York: Guilford Press.
- Peuquet, Donna J. 1994. It's about time: A conceptual framework for the representation of temporal dynamics in geographic information systems. *Annals of the Association of American Geographers* 84 (3):441-61.
- Rumsey, David. 1996-. *David Rumsey Map Collection* <http://www.davidrumsey.com/> (last accessed 18 February 2013). Cartography Associates.
- Social Explorer*. Led by Andrew Beveridge. <http://www.socialexplorer.com> (last accessed 18 February 2013).
- Southall, Humphrey. 2003. *Great Britain Historical Gazetteer/GIS*. University of Portsmouth: Great Britain Historical Gazetteer/GIS.
- Southall, Humphrey, P. Manning, M. Berman, J. Gerring, and P. Bol. 2011. *Understanding global change: how best to organize information?* University of Portsmouth's Research Repository.
- Vakgroep nieuwste geschiedenis/Department of Modern History at Ghent University. *Belgian historical GIS*. <http://www.lokstat.ugent.be/> (last accessed 18 February 2013).
- Warf, Barney, and Santa Arias. 2009. *The spatial turn: interdisciplinary perspectives*. *Routledge studies in human geography*. London; New York: Routledge.

- Welshons, Marlo, et al. 2006. *Our cultural commonwealth: the report of the american council of learned societies commission on cyberinfrastructure for the humanities and social sciences*. New York: American Council of Learned Societies.
- White, Richard. 2010. What is Spatial History? Stanford University: Spatial History Lab.
- . 2011. *Railroaded: the transcontinentals and the making of modern America*. New York: Norton.
- Wick, Mark. 2012. Geonames. <http://www.geonames.org> (last accessed 18 February 2013).
- Withers, Charles W. J. 2009. Place and the "spatial turn" in geography and history. *Journal of the History of Ideas* 70 (4):637-58.
- World-Historical Dataverse <http://www.dataverse.pitt.edu/> (last accessed 18 February 2013). 2011-. Center for Historical Information and Analysis, University of Pittsburgh.
- Yost, A.Y., and W.J. Jr. Carswell. 2009. Geographic names: U.S. geological survey fact sheet 2009-3016. <http://pubs.usgs.gov/fs/2009/3016/> (last accessed 18 February 2013).
- Correspondence:* Center for Geographic Analysis, Harvard University, 1737 Cambridge St., Cambridge MA 02138, email: pkbol@fas.harvard.edu.