On an Infrastructure for Historical Spatial Analysis

for Perspectives on History/AHA

On an infrastructure for historical spatial analysis

Peter K. Bol
Harvard University, Center for Geographic Analysis
pkbol@fas.harvard.edu

References to a "spatial" turn have been increasing across the disciplines (Guldi 2011; Warf and Arias 2009). A decade ago the geographer Anne Knowles connected a spatial turn in history to historical geographic information systems (GIS) in an issue of Social Science History (Knowles 2000). She followed this with collections illustrating the application of GIS to historical subjects (Knowles 2002; Knowles and Hillier 2008).

Geospatial analysis links geography and history, temporal change and spatial variation, in a more technological and data driven fashion than the place-centric research typical of the "geo-humanities" (as in Dear et al. 2011). A number of major historical studies using geospatial analysis have appeared in the last decade (Ayers and Rubin 2000; Ayers 2003; Gordon 2008; White 2011).

History and geography have been, and remain, divided, as the subtitle of Alan Baker's *Geography and History*: *Bridging the Divide* (Baker 2003) recognizes. Still, there is a great deal of interest on the part of the geographers—the Association of American Geographers had sixty sessions on "Space-Time Integration in Geography and GIScience" at its 2011 meeting. A tool such as GIS helps the historian see what was not evident before. It enables the use of large quantities of data, the visualization and measurement of the spatial patterns in that data, the tracking of change over time, and the correlation of those patterns with information from different domains, Historical GIS enables historians to take advantage of some of the accomplishments of GIScience.

Learning to apply GIS to historical questions is not more demanding than learning to use GIS software (which does have a steep learning curve); there is a useful textbook for historical GIS (Gregory and Ell 2007) and many more for GIS software. Many institutions have software licensing for the most sophisticated GIS package, ArcGIS. There is also useful freeware such as QuantumGIS which works on both a PC and a MAC. Beyond the time necessary for learning software, the larger obstacles to making geospatial analysis useful for historians are infrastructural. In what follows I note some of some approaches to the major cyberinfrastructural challenges.

1. An Adequate World-historical Gazetteer.

GIS is about geographic space, but in the historical record "place" matters more than "space." A person is of a place, a religious site is located at a place, tax is reported by place, a postal station is a place in itself. Places are nodes in networks (for premodern times it is easier to find the nodes than the routes between them and reliable sources for administrative boundaries before 1800 are few). 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 – creates an intelligible interface between the geophysical world and human culture; the name makes it a place. If we can capture names in written/drawn sources and locate them in space, then we can locate historical data with spatial attributes (tax records, population, religious activities, battles) in space. Naming pertains to all aspects of human life and, like everything in human culture, names are not stable. A name links the historical and the geographic but its validity is temporally bounded. The historical record is, ultimately, finite, so in principle we can 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 in most cases 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 major online world gazetteers are invaluable but share a common flaw: they ignore time. This is true of the Geographic Names Information System (http://geonames.usgs.gov/), which provides over 2 million names for natural and constructed places (except roads) in the United States (Yost and Carswell 2009); the National Geospatial-Intelligence Agency's GEOnet Names Server (http://geonames.nga.mil/ggmagaz/), the official repository for foreign place names, which has over 5 million foreign geographic features, including alternate names and the local vernacular (National Geospatial-Intelligence Agency 2012); and GeoNames (http://www.geonames.org), the greatest non-governmental gazetteer (founded by Mark Wick), which contains over 10 million geographical names including 7.5 million unique features, including 2.8 million populated places and 5.5 million alternate names (Wick 2012). GeoNames handles 20 million daily web service requests.

The US federally mandated gazetteers will not include time unless that is part of their legal charge. This 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 texts, archiving past records and thus creating a written trail. When an information management system keeps up-to-date by overwriting earlier data that information system is sacrificing a longitudinal record to clerical efficiency. Thus a first-order infrastructural need in linking history and geography, time and space, is a world historical gazetteer.

A world-historical gazetteer is fundamental to spatially-conscious historical research; the planning process to define the data models, content standards and interoperability protocols for multilingual online gazetteer system is in place but the funding for building it is not. A number of historical GIS databases provide the kinds of information that a world-historical 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 (Southall 2003), accessible through the A Vision of Britain through Time website (*http://www.visionofbritain.org.uk/*). The 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 (*https://www.nhgis.org/*) 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 print *Historical Gazetteer of the United States* (Hellmann 2005). In contrast the China Historical GIS (www.fas.harvard.edu/~**chgis**), covering 221 BCE-1911 CE, was created as a time series of administrative entities and major towns and their changing administrative relationships. 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). The AAG's Historical GIS Clearing House and Forum provides a listing of historical GIS projects and gazetteers (http://www.aag.org/cs/projects_and_programs/historical_gis_clearinghouse).

Computationally mining existing databases, dated digitized texts, and digitized georeferenced and dated maps could populate a world-historical gazetteer. Beginning in the 1790s mathematically accurate cartography has mapped the globe. These maps provide information on routes, boundaries, physical features, and locations that texts cannot provide. For a limited historical period—but one which saw global 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 (Chiang and Knoblock). Until such a project is funded the best hope may lie with projects now beginning for crowd sourcing the extraction of data from scanned maps.

2. Access to Georeferenced Maps and Spatial Data

Spatial analysis uses both geo-referenced scans of historical maps ("raster data") and spatial datasets that attach information to points, line, and polygons ("vector data"). Map libraries, such as Harvard, Berkeley, the University of Texas, and the Library of Congress, usually make historical maps they have scanned available. The David Rumsey Map Collection (http://www.davidrumsey.com) has led the way, it makes 29,000 scanned maps available online of which some 22,000 have rough geo-referencing.

A federated system for registering and sharing historical map scans now exists through Old Maps Online (http://www.oldmapsonline.org), a UK-based project. In effect it creates a union catalog of georefernced historical maps, including the Rumsey collection, the New York Public Library, and soon Harvard and the Boston Public. The user interface allows for searching on a map of the world, narrowing the search by time and space.

OpenGeoportal.org (http://opengeoportal.org/), led by Tufts, Harvard and MIT with a growing number of partner institutions, has developed a common portal for their digitized holdings (soon 40,000 layers) of vector and raster data, scanned maps, and aerial photography. The OGP will incorporate cloud-based automated metadata harvesting; robust web mapping via GeoCommons, ArcGIS Online, Worldmap, and Ushahidi; geospatial thesaurus for rapid data discovery, and new visualization tools to discover areas of high and low data coverage.

A concomitant to a federated map/spatial dataset catalog is a system for archiving and searching historical datasets that could be joined to spatial objects in a GIS. The World-Historical Dataverse of the Center for Historical Information and Analysis at the University of Pittsburg is creating a system the collection, preservation and analysis of world historical data (http://www.dataverse.pitt.edu).

Some large projects that promise to make large amounts of data available. Terra Populus, at the University of Minnesota, has been funded to combine two centuries of census data with global environmental data including land cover, land use and climate records and disseminate the results to the world (http://www.pop.umn.edu/terra_pop). In contrast, there are valuable projects that allow interactive data query visualization online, but are not necessarily geared to making their datasets available for user download. Examples include the Republic of Letters visualization of early modern letter exchanges (https://republicofletters.stanford.edu/) and Orbis, which tracks the time and cost of transportation and communication in the Roman empire (http://orbis.stanford.edu/), both from Stanford, and the *Digital Atlas of Roman and Medieval Civilization* (http://darmc.harvard.edu/) from Harvard. The Social Explorer (http://www.socialexplorer.com) platform has free and subscription editions for the visualization of spatialized data (such as historical US census data). MapStory aims to bring time and space together online. Users may upload their own GIS data with time attributes, display them dynamically on the online map, and allow others to download the data (http://mapstory.org/).

3. Online Platforms for Accessing and Sharing Spatial Data

The final piece of cyberinfrastructure is an online platform for sharing spatialized data (both online visualization and download). There has been significant progress. ESRI's ArcGIS Online (http://www.arcgis.com/home/) and ArcGIS

Explorer Online (http://explorer.arcgis.com/) allow users to create, store and share maps and datasets, manage geospatial content, and control access to volunteered content. Geocommons (http://geocommons.com/) developed by the GeoIQ company (http://www.geoiq.com) is similar. In collaboration with the open-source web-mapping developer community the Center for Geographic Analysis at Harvard is developing the WorldMap platform (http://worldmap.harvard.edu/) allowing users to explore, visualize, edit 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 (http://warp.worldmap.harvard.edu), and publish data to a few collaborators or to the world. The promise of WorldMap is that it is cumulative, nearly four-thousand registered users are already uploading and sharing their data and maps with over 140,000 viewers. Working with MapStory it will soon add time animation, bringing together geography and history.

Andrienko, G., N. Andrienko, U. Demsar, D. Dransch, J. Dykes, S.I. Fabrikant, M. ern, M.J. Kraak, and [et al.]. 2010. Space, time and visual analytics. *International journal of geographical information science* 24 1577-1600.

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*, edited by A. K. Knowles and A. Hillier. Redlands, CA: ESRI Press.

Bol, Peter K., and Jianxong Ge. 2005. China Historical GIS. *Historical Geography* 33 150-2.

Chiang, Yao-Yi, and Craig A. Knoblock. Recognition of Multi-Oriented, Multi-Sized, and Curved Text. In *Proceedings of the Tenth International Conference on Document Analysis and Recognition*.

Dear, Michael, Jim Ketchum, Sarah Luria, and Doug Richardson, eds. 2011. *Geohumanities: Art, History, Text at the Edge of Place*. London and New York: Routledge.

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.

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 [cited January 16, 2012. Available from http://spatial.scholarslab.org/spatial-turn/disciplinary-perspectives/the-spatial-turn-in-history/.

Hellmann, Paul T. 2005. *Historical gazetteer of the United States*. New York: Routledge.

Knowles, Anne Kelly, ed. 2002. *Past time, past place : GIS for history*. Redlands, Calif.: ESRI Press.

Knowles, Anne Kelly ed. 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.

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.

Southall, Humphrey. 2003. Great Britain Historical Gazetteer/GIS. University of Portsmouth: Great Britain Historical Gazetteer/GIS.

Warf, Barney, and Santa Arias. 2009. *The spatial turn : interdisciplinary perspectives*, *Routledge studies in human geography*. London ; New York: Routledge.

White, Richard. 2011. *Railroaded : the transcontinentals and the making of modern America*. 1st ed. New York: Norton.

Yost, A.Y., and W.J. Jr. Carswell. 2009. Geographic Names: U.S. Geological Survey Fact Sheet 2009-3016.