Historical GIS Projects: Spatial Data Infrastructure

Robin M. N. Parker

Abstract: The use of historical GIS (HGIS) in humanities and social sciences research has added dimensions to scholarship in terms of both analysis and visualization. The construction of appropriate HGIS systems for the integration of historical data requires large investments in time, resources, and technical expertise. Fundamental to the success of such systems is the spatial data infrastructure (SDI) that consists of crucial components including licensing, data formats, documentation, and standards of metadata. This paper examines the aspects of an SDI necessary for HGIS, particularly on the level of national endeavours, through use of the example of the Great Britain Historical GIS Project. The detailed facets of an effective SDI for a national HGIS can serve as a model for researchers in Canada interested in developing a similar resource.

About the Author(s): Robin M. N. Parker is currently a Master of Library and Information Studies candidate in her second year. Following her BSc in History and General Science, she strives to integrate her interests in current medical information services and historical studies. This paper was originally written for a School of Information Management course, Geospatial Information Management and provided background information for Robin's research activities with Dr. Fiona Black regarding HGIS applications in the history of print culture.

Introduction

Historical geographic information systems (HGIS) have been increasingly popular additions to scholarly research in a range of historical endeavours over the past decade. HGIS take the analytical capacity of databases holding historical data and displays the resulting interpretation using visual representations of the spatial information. A general definition of geographic information systems is given by Michael Goodman (2000), who states:

GIS is defined most generally as technology for processing a specific class of information—geographic information. *Processing* is understood to encompass creation, acquisition, storage, editing, transformation, analysis, visualization, sharing, and any other functions amenable to execution in a digital domain. *Geographic information* is readily defined as information linking locations on the Earth's surface with specific properties, such as name, feature, population, elevation, temperature. (p. 159)

In the interest of building the capacity to conduct historical analyses using this form of processing and visualization, many projects have been established to create the necessary framework. Notable national HGIS efforts include the National Historic GIS (NHGIS) project in the United States, the Great Britain Historical GIS (GBHGIS) project in the UK, and the China Historical GIS (CHGIS) project. Equally notable is the absence of a unified effort in Canada to build the infrastructure to support research using historical GIS.

In the final chapter of their 2007 book, Ian Gregory and Paul Ell discuss the benefits of building a national historical GIS infrastructure. They point out that the large investment of time and resources required for such an endeavour provides benefits in the long run for small and medium-sized research projects, because they have easy access to the data required and can proceed directly to analyses. Meanwhile Fiona Black has pointed out the value of and preferences for geographic visualization for print culture historians in an upcoming publication. Her research into the merits of HGIS for the history of the book makes a case that can be extrapolated to other historical research.

Using the existing projects as models, in particular focusing on the GBHGIS, the purpose of this paper is to describe the components needed to bring Canada's HGIS infrastructure to a similar level as its international counterparts. In order for Canadian scholars to be able to employ HGIS in their research without individually replicating the time-consuming processes inherent to building the basic components, a coordinated effort to construct a spatial data infrastructure for historic information is necessary.

Spatial data infrastructure

The term spatial data infrastructure (SDI) refers to the "collection of technologies, policies and institutional arrangements that facilitate the availability of and access to spatial data" (Global Spatial Data Infrastructure Association, 2004, p. 8). Furthermore, the purpose of the SDI is to

facilitate the "spatial data discovery, evaluation, and application for users and providers within all levels of government, the commercial sector, the non-profit sector, academia and by citizens in general" (Global Spatial Data Infrastructure Association, 2004, p. 8). This paper will examine these characteristics of a historical GIS by looking at the example of the GBHGIS Project. The project has allowed access to a substantial collection of historical shape boundaries and statistical information. The current version of the project that is publically accessible on the web is not a true GIS: "[a]lthough th[e] new system is maybe more of a database than a GIS, it is the geographical information in it which makes it unique – so we now call the whole system the "Great Britain Historical GIS", not just the collection of digital boundaries" (University of Portsmouth, 2009b). The process to build such a system and the possible applications thereof are similar to those of other geographic-based systems requiring a SDI.

Challenges for historical GIS construction

Historical GIS has similar data infrastructure needs as other GIS projects, but also has the additional complication of changes over time The temporal aspect of HGIS is a challenge to traditional GIS systems and requires linkage to extra databases or tables that can be queried to reflect changes to the spatial data, as well as associated attributes such as place names, social trends, and industrial developments. As Goodchild has stated, "[e]arly GIS database designs created great difficulty for this type of application because the map metaphor implied that any layer would be a single snapshot, making it impossible to organize multiple snapshots into a coherent database" (2008, p. 194). While this issue has not been completely resolved, techniques have been developed to deal with the problems presented by the temporal shifting of the boundaries that are frozen in the form of boundary files.

National HGIS projects

In 2005, Anne Knowles edited an article compiling descriptions of national historical GIS projects. The resulting summaries of HGIS efforts provide an overview of achievements at that time. For example, the American venture, called the National Historical GIS (NHGIS), developed and hosted by the University of Minnesota, created a resource that "consists of three major components: data and documentation, mapping, and data access" (p.135). As will be discussed later, the latter component consisted of online, public access to the other parts of the project and used the same DDI metadata standards as the GBHGIS. The project to construct a national HGIS for China is of similar scope to the GBHGIS and NHGIS and was coordinated and hosted at Harvard University. Peter Bol (2007), a key member of the CHGIS team, has written elsewhere on the process and decisions concerned with that endeavour. Other projects in the Knowles review vary in scope and comprehensiveness, and include the national projects of Ireland, Belgium, the Netherlands, Germany, Russia, and Korea.

Canada

Canada, on the other hand, has not undertaken a comprehensive national project for HGIS. The project documented by Knowles (2005) was begun by Fiona Black and Bertrum MacDonald in their study of the spatial aspects of the history of print culture. However, that project consists mainly of a text-based database that includes spatially-referenced records, but is not linked to a large-scale, publically accessible GIS platform. These scholars, with backgrounds in library and information studies, have pursued analysing the Canada book trade data with the capabilities of a HGIS, but have been hampered by the substantial nature of the endeavour and the lack of a national project upon which to build their research.

The relatively narrow scope of the proposed history of print culture project also differentiates it from those such as the NHGIS, GBHGIS, and the CHGIS, which are wider in scope and therefore have a more inclusive infrastructure. The Canadian project at this point constitutes a research-driven HGIS. Bol differentiates such a project from an infrastructural HGIS:

> The distinction between a research-driven GIS versus an infrastructural GIS is similar to the difference between the material one borrows from a library for a specific research project and the library itself. An infrastructural HGIS is thus like a library: it aims to serve many different research questions. (2007, p. 4)

Therefore, the overall spatial data infrastructure ultimately needed in Canada is similar and the creation of a complete national project with similar components as the British or American counterparts would provide the necessary resources for this and many other research projects. Furthermore, the use of historical GIS in the UK to demonstrate precedents and lessons from the past that could influence current policies and actions would also be useful in Canada. In particular, using HGIS to examine historic natural resources usages and the causal societal interactions could help inform choices that the citizens and governments in Canada make for the future.

Canada now has the opportunity to learn from the experiences of other national HGIS projects and benefit from the progress in project design and technical developments. Using the GBHGIS—in combination with a comparison of other national projects—as a model, could facilitate the creation of a national Canadian HGIS. Thus Canada's social science researchers would have a tool to rival our international colleagues and contribute to the overall accomplishment of a global HGIS, as will be discussed in the conclusion of this article.

Great Britain Historical GIS Project

History of GBHGIS

In the case of the GBHGIS project, efforts have been underway since 1994 to develop a coordinated database of information available for use by UK scholars. The database that

originated the project was not geographically based but has been expanded and linked to spatial information, and later a digital map collection, in order to gradually build a hallmark HGIS. The project has undergone a number of funding developments from various organizations and been based out of a series of academic institutions. However, there has been uniformity in the core staff and directorship in the program since its early days. Because of the immense effort involved in the development of an HGIS structure, the "soft system" of the personnel committed to the project is as important as any of the technology or metadata around which the system is constructed. Bol (2007) also mentions the need for a committed team to ensure success for the CHGIS project.

Humphrey Southall and Ian Gregory are two of the primary individuals involved with the GBHGIS. Southall was originally involved with the Labour Markets Database, from which the later incarnations of the project developed (University of Portsmouth, 2009b). The seminal database was based on statistical information and was subsequently linked to boundary mapping, under the direction of GIS-trained Gregory. Gregory's background in geography was developed through undergraduate and graduate studies and his involvement with the GBHGIS provided the basis for his doctoral thesis (University of Lancaster, n.d.). Many other developers and scholars were also involved in the evolution of the project, with constant consultation on various mailing lists concerning the technical aspects of the platforms.

Most recently, with a re-launch in the summer of 2009, the GBHGIS has changed to allow for the creation of an interactive public access website, the Vision of Britain through Time portal, and has added an expansive collection of digital maps to the project. This evolution has changed the structure of the project, as will be discussed in the next section and greatly increased the number of users that can now access and explore the resource, making it a true national project.

Structure

The GBHGIS project links the various statistical data sources, boundary information, and a digital library of historical maps. A schematic of the current structure by which a complex array of information is pulled together as follows:
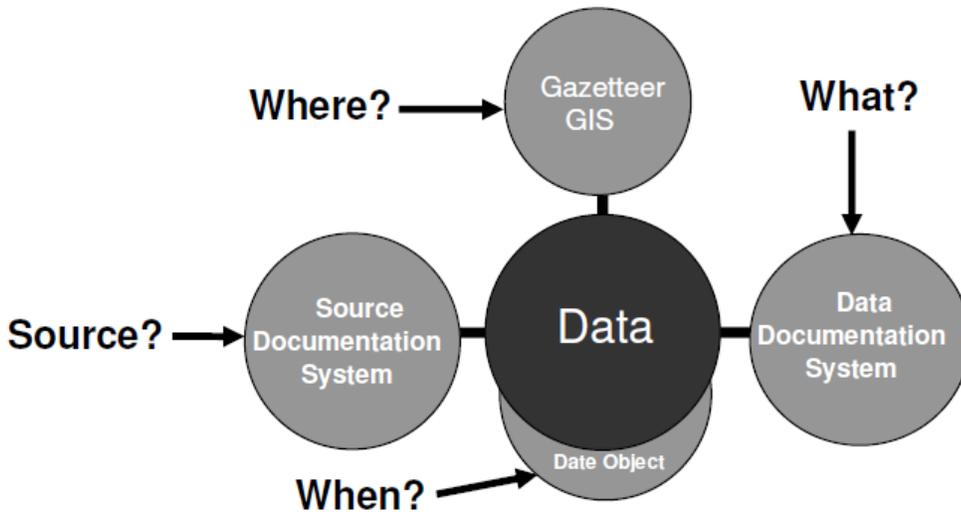
GBHGIS Data Architecture

Source: Southall, H. (2009)

The use of a gazetteer as a type of joining table to act as an authority list for place names allows a linkage between the standard names that appear in the GIS database and the multiple name variations that may appear in the repository of the statistical data. The variant names may be a reflection of changes in spelling over time or discrepancies coming from different sources. This is one example of the use of an ontology, or structured hierarchy, in order to connect the divergent parts of the entire system, which is necessary to make use of the spatial data stored in the GIS or mapping component of the HGIS.

The function of the gazetteer is illustrated below.

Format

Through the incarnations of the system, the data, both spatial and statistical, has been stored in a variety of formats. Currently, there are two versions maintained. The first is the Oracle database that holds the assorted tables in their relational configuration. On the other hand, the web-based component has been maintained on open source software (OSS) stores the statistical data in a single table, wherein the other columns include the descriptive and spatial information pertaining to each entry (University of Portsmouth, 2009a; Great Britain Historical GIS Project, 2009). The web page describing data documentation for the Vision of Britain project is "structured as an ontology containing 6,152 entities of [many] types, linked by 19,427 relationships" (Great Britain Historical GIS Project, 2009). The OSS, Postgres, and its GIS equivalent, PostGIS, hold the statistical and boundary data respectively. In addition, the web content uses another open source platform, MapServer, to connect to a digital library of scanned and georeferenced maps.

Types of data

The GBHGIS contains historical statistical data retrieved from a variety of sources, including censuses and election results, as well as spatial data in the form of boundary polygon shape files. In addition, the expansion of the project with the most recent round of substantial funding allowed for the digitization of hundreds of geo-referenced historical maps. This digital library is linked to the Vision of Britain component of the system in order to allow scholars and other users to better visualize the places and time periods of interest. This particular project is also unique in the inclusion of a substantial collection of textual information, drawing from historical travel writing, descriptive gazetteers, and censuses. References to places in these records are then linked to the rest of the system in order to enrich the levels of relationships possible.

Data documentation/metadata

The documentation of the data, the sources, and the processes has been well maintained in the process of developing the GBHGIS. The project plan for the most recent stage of the system, accomplished with funding from the Joint Information Systems Committee, notes the various aspects of project documentation including memorandums of understanding with partner organizations, licences, and technical documentation (University of Portsmouth, 2007).

The same document notes the existing use of the metadata standards from the Digital Documentation Initiative (DDI) for the data compiled from statistical sources. These are the same documentation standards that are used for the National Historical GIS project undertaken in the United States (Knowles, 2005). Many of the objects in the database are supported by metadata in Dublin Core format. Mapping standards from the Open Geospatial Consortium were used for creating digitized maps for the web (University of Portsmouth, 2009a).

In Canada, the existing Canadian Geospatial Data Infrastructure endorsed standards include data visualization with Web Map Service, encoding and transfer with Geographic Markup Language, and data access with Web Feature Service (GeoConnnections, 2008a). Any national HGIS project should adopt these standards or compatible ones in order to assure cross-functionality of data across platforms.

Size

While none of the documentation available indicates the size of the data stored, the description of the project covers the number of records and spatial units included in the GBHGIS. The general information page on the project website states that there are 48,000 units with over 150,000 relationships. Furthermore, the spatial component contains more than 40,000 boundary polygons associated with those units, many of which are assigned dates to indicate changes over time (University of Portsmouth, 2009b).

The author's experience with a research-driven HGIS project for the study of the book trade in Canada suggests that the data storage needs involved are considerable. The digital library of scanned historical maps constitutes the most substantial data size component of a HGIS that uses historical sources for visualization. Additionally the database of textual records that links to the spatial information requires adequate storage and operating capacity to perform the necessary analyses.

Rationale/purpose

The main purpose of the GBHGIS is to facilitate social science research that takes the temporal-spatial aspects into consideration. As the creators state on the website about the project, "[t]he central focus of the project is on the development of our GIS as a national resource. However, there have been a series of associated major research projects" (University of Portsmouth, 2009a). In a paper written in 2002, about mid-way through the project development, Gregory and Southall describe the goal of the endeavour:

> to build a flexible net of administrative unit boundaries that could capture the fluid reality of society as accurately as possible for any date and for any of several kinds of units, giving social science historians a better alternative to the rigid, often anachronistic frames available in printed maps. (p. 120)

User needs

In order to use the Vision of Britain portal, the user only needs Internet access. However, those researchers affiliated with UK universities can get access to the password-protected datasets held at Edinburgh University and Essex University (University of Portsmouth, 2009d). In order to use the historical boundaries and statistics, held respectively at the two universities,

researchers need to integrate the downloaded datasets into a GIS platform, to allow further manipulation and analyses.

Access/licensing

The funding for the GBHGIS has largely come from public sources such as tax-subsidized government programs and the National Lottery, therefore part of the mandate is to allow for public access. This has been achieved with the Vision of Britain component of the project. However, as previously mentioned, this part of the GBHGIS in particular is more of a database than a GIS because it does not allow downloading and manipulation of the actual spatial datasets. This fact is due in part to the restrictions surrounding the project coming from the source of the funding. On the other hand, the datasets that constitute the GBHGIS are available to academics in the UK. This scenario of tiered access differs from that of the NHGIS in the United States, where the entire datasets are freely available to the public.

Use scenarios

The website lists a number of examples using the GBHGIS, including using the GIS to examine the pre-history of the north-south divide in Britain, the changing geography of health, the London GIS, and early mapping from Domesday to the first census (University of Portsmouth, 2009c). Other examples mentioned are "an analysis of the impact of high unemployment in inter-war Britain on both short-run infant mortality rates and long-run life expectancy" (University of Portsmouth, 2009b, conclusion).

Meanwhile, in Canada, the possible use scenario that has pushed HGIS the furthest is the history of print culture. While lacking a national infrastructure such as the GBHGIS, scholars at Dalhousie University have conceptualized the usefulness of HGIS in the study of the history of the book trade in Canada. The figure below demonstrates one way that disparate information pertaining to the book production, trade, and use could be layered and simultaneously queried using a database that then re-projects the information in a spatial format.
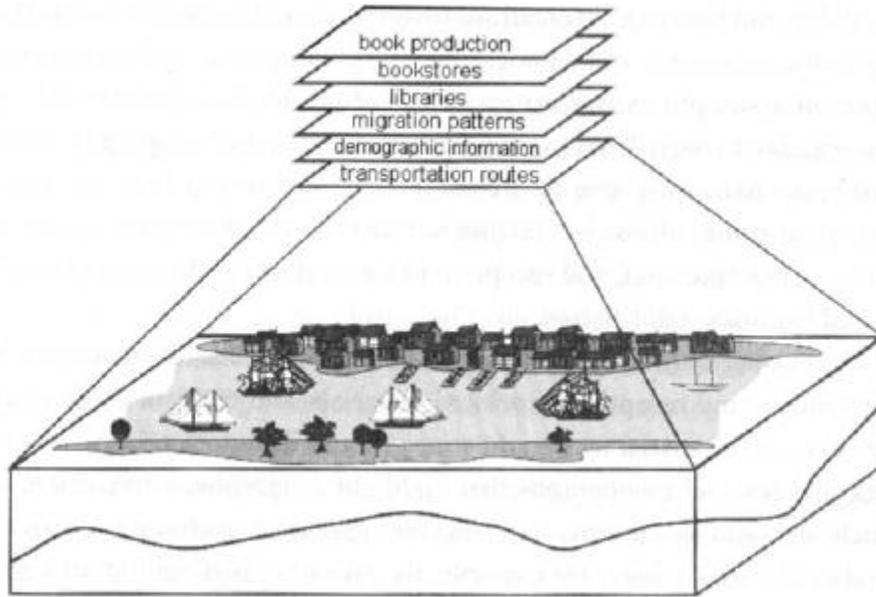
[Abstracting the real world into layers for the study of book history]

Source: MacDonald and Black, 2000 (as cited in Gregory, 2005)

In the United States, the Spatial History Project at Stanford University has pursued a number of use scenarios, including examinations of the shaping of the West through railroads, geographies of the Holocaust, and explorations of Chile's aquaculture industry, amongst many others (Stanford University, 2010). These investigations have undoubtedly benefited from the infrastructure provided by the NHGIS, particularly the freely available datasets of historical boundaries.

Cost and funding

The overall cost of the GBHGIS is estimated at approximately £2.1 million since the official beginning of the project in 1994 (University of Portsmouth, 2009a). These funds have come from an assortment of grants from such organizations as the Economic and Social Research Council, the Wellcome Trust, the Joint Information Systems Committee, and most substantially, the UK National Lottery. The latter two have in particular contributed to the public access parts of the Vision of Britain project. All together, at least two dozen organizations have supported the project over the span of its development. Not least of these would be the participation of numerous academics and their home institutions.

In another example, the CHGIS project for China was accomplished with nearly $1.5 million (USD) from a variety of sources. Bol describes gathering funding through grants from the Luce Foundation, the National Foundation for the Humanities, and matching funds from two Harvard

institutes (2007). The comprehensive nature of both the CHGIS and the GBHGIS explain the substantial funding requirements needed to realize the projects.

Ethics and Privacy

Due to the historical nature of the information, most of the ethical and privacy issues are less relevant to HGIS than some other GIS applications. The statistical information that is publically available for more recent censuses is compiled and summarized in such a way to minimize the negative consequences to individual or group privacy. Other information, such as the boundaries and other spatial data, would be within the public domain and therefore has limited ethical considerations. As with any research, there is the potential of ethical issues arising from the way the data and subsequent results are used. However, these issues are not unique to HGIS-based research and are addressed by research ethics boards.

Many of the same privacy issues that pertain to any historical or genealogical research also apply to HGIS projects and the issues are in fact amplified by the spatially referenced nature of the resulting analyses. The sensitive information that may be uncovered through historical research is easier to associate with identifiers such as location through HGIS. In addition, the wider dissemination and ease of data manipulation through web-based HGIS interfaces could increase the significance of privacy and ethical concerns.

As with other forms of historical investigation, the limitations of the sources must always be acknowledged in order to minimize the risk of misinterpretation. For example, statistical representations through HGIS in which some of the data has been lost or is unavailable could contribute to misleading conclusions if the omissions are not noted and taken into account. Researchers must be explicit regarding the context of both the data and the resulting interpretations. As Gregory and Healey (2007) describe, the ambiguity of the historical data, both textual and spatial, should be accounted for through the documentation of the HGIS project. In other words, the metadata should include descriptions of the data and its limitations so that the onus is on the individual researchers to use the data and the resulting analyses responsibly.

Policy

As previously discussed, the limited access of the GBHGIS datasets to academics is due to the funding that made the Vision of Britain website possible, as well as the Crown ownership of the original data from the Ordnance Survey. Other national projects such as the American counterpart allow free access to all of the datasets maintained by the project. In terms of a potential Canadian project, the model of open access—for both the datasets and the system platform—would be ideal. Such a policy of open access would be in keeping with the present policies regarding other types of Canadian geospatial data. Currently geospatial data is available through initiatives such as GeoConnections in both no-fee and fee-based distribution models (GeoConnections, 2008b). This data is subject to a number of relevant Canadian

policies concerning information management, information technology, and access to information.

Digital Earth historical component

Ideally, the various national historical GIS projects could all contribute to the historical perspectives of the Digital Earth vision. The Digital Earth concept is succinctly described on the website for the Fifth International Symposium on Digital Earth as

> the virtual and 3-D representation of the Earth that is spatially referenced and interconnected with digital knowledge archives from around the planet [...] of scientific, natural, and cultural information to describe and understand the Earth, its systems, and human activities." (2007, para. 1)

Thus, a complete Digital Earth project would necessarily include historical data and spatial attributes. Linking national HGIS endeavours to cover the globe would provide a framework to examine temporal aspects of this spatial representation.

The necessary groundwork for implementation of this type of overlay includes the compatibility of the various systems. Interoperability of the standards used for both the data and the mapping components of the HGIS is required. The universal application of the DDI standards adopted by both the GBHGIS and the NHGIS will facilitate the possibility of such a global HGIS and allow linkages throughout the system. Further use of Ethnologue and Linguist, the language standards employed by the GBHGIS, would allow the historical Digital Earth project to recognize and connect place names and possibly text across cultural boundaries. In addition, the use of one set of standards, or assorted compatible ones, for mapping, will be necessary. The work of the Open Geospatial Consortium to establish and promote such standards would contribute to the success of such an endeavour.

The creation of a Canadian HGIS would add our unique cultural and geographic perspective to the cumulative historical component of a Digital Earth project. However, even before such an end result is achieved, historical scholars across Canada would benefit from the additional analyses a national historical GIS could contribute to their research. The Canadian national project could be built using the lessons learned from other equivalent projects such as the GBHGIS, the NHGIS, and the CHGIS. These earlier projects would serve as a composite model to which a Canadian approach should be used to confront the distinctive challenges of the Canadian experience and landscape.

References

Black, F.A. (2010, forthcoming). Construing the spaces of print culture: Book historians' visualization preferences. In M. Ogborn & C. Withers (Eds.), *Geographies of the Book*. Ashgate Press: Surrey, UK.

Bol, P. (2007). The China Historical Geographic Information System (CHGIS): Choices faced, lessons learned. Paper presented at the Conference on Historical Maps and GIS, Nagoya University, August 23-24, 2007. Retrieved from http://www.fas.harvard.edu/~chgis/work/docs/papers/BOL_CHGIS_Lessons_Learned.pdf

Fifth International Symposium on Digital Earth. (2007). *About Digital Earth*. Retrieved from http://www.isde5.org/about_digitalearth.htm

GeoConnections. (2008a). Building and sustaining the Canadian Geospatial Data Infrastructure. Retrieved from http://www.geoconnections.org/publications/General_information/developers_english.pdf

GeoConnections. (2008b). The dissemination of government geographic data in Canada: Guide to best practices, version 2. Retrieved from http://www.geoconnections.org/en/resourcelibrary/keyStudiesReports

Global Spatial Data Infrastructure Association. (2004). *Spatial Data Infrastructure cookbook v2.0* [PDF]. Retrieved from http://www.gsdi.org/docs2004/Cookbook/cookbookV2.0.pdf

Goodchild, M.F. (2000). New horizons for the social sciences: geographical information systems. *Canadian Journal of Policy Research, 1*(1), 158 – 161.

Goodchild, M.F. (2008). Combining space and time: new potential for temporal GIS. In A. K. Knowles (Ed.), *Placing history: How maps, spatial data, and GIS are changing historical scholarship* (pp. 179-197). Redlands, CA: ESRI Press.

Great Britain Historical GIS Project. (2009). *A vision of Britain through time: Data documentation.* Retrieved from http://visionofbritain.org.uk/data/

Gregory, I.N. (2005). *A place in history: A guide to using GIS in historical research* (2nd ed.). Belfast, Northern Ireland: ESRC's Research Methods Programme. Retrieved from http://www.lancs.ac.uk/staff/gregoryi/

Gregory, I.N. & Ell, P.S. (2007). *Historical GIS: Technologies, methodologies and scholarship*. Cambridge, UK: Cambridge University Press.

Gregory, I.N. & Healey, R.G. (2007). Historical GIS: structuring, mapping and analysing geographies of the past. *Progress in Human Geography 31*(5), 638-653. doi: 10.1177/0309132507081495

Gregory, I.N. & Southall, H.R.. (2002) Mapping British population history. In A. K. Knowles (ed.) *Past time, past place: GIS for h* (pp. 117-130). Redlands, CA: ESRI Press.

Knowles, A.K. (Ed.). (2005). Report on national historical GIS projects. *Historical Geographer, 33*, 134-58.

Southall, H. (2009). GB Historical GIS data architecture: A place for everything and everything in its place. Retrieved from http://www.port.ac.uk/research/gbhgis/mediaresources/freearticles/

Stanford University. (2010). *Spatial history project.* Retrieved from http://www.stanford.edu/group/spatialhistory/cgi-bin/site/index.php

University of Lancaster. (n.d.) *Ian Gregory: Digital humanities.* Retrieved from http://www.lancs.ac.uk/staff/gregoryi/

University of Portsmouth. (2007). *HBB Project Plan* [PDF]. Retrieved from http://www.port.ac.uk/research/gbhgis/abouttheproject/currentactivities/historicboundariesof britain/filetodownload,72500,en.pdf

University of Portsmouth. (2009a). *Great Britain Historical Geographical Information System (GBHGIS) About the project.* Retrieved from http://www.port.ac.uk/research/gbhgis/aboutthegbhistoricalgis/

University of Portsmouth. (2009b). *Great Britain Historical Geographical Information System (GBHGIS) History.* Retrieved from http://www.port.ac.uk/research/gbhgis/abouttheproject/history/

University of Portsmouth. (2009c). *Great Britain Historical Geographical Information System (GBHGIS) map samplers.* Retrieved from http://www.port.ac.uk/research/gbhgis/aboutthegbhistoricalgis/mapsamplers/

University of Portsmouth. (2009d). *Great Britain Historical Geographical Information System (GBHGIS) online access.* Retrieved from http://www.port.ac.uk/research/gbhgis/aboutthegbhistoricalgis/onlineaccess/