Constructing a Vision of Britain through Time:
Integrating old maps, census reports, travel writing, and much else, into an online historical atlas

Humphrey Southall (University of Portsmouth/Great Britain Historical GIS)

Introduction

Geographical Information Systems (GIS) technology has a long history of applications in the humanities, but many of these applications concern things with a physical existence in the landscape, and so with archaeology. This paper is a description of the Great Britain Historical GIS (GBH GIS), and its associated web site A Vision of Britain through Time (VoB), and these are historical in that they concern and are based on the written records of the past: at heart, the GBH GIS is a compendium of geographical surveys of Britain. This we argue benefits from an approach significantly different from traditional GIS, geo-semantic rather than geo-spatial, and concerned as much with named places as with geometrical spaces. Some other key features should be noted:

Firstly, our main focus is the relatively recent past, and in particular the period since 1801 when Britain carried out its first census, so much of our content is statistical. One consequence is that it is more useful to compare our system with the US National Historical GIS than with systems developed by archaeologists. However, we discuss below the addition of books written by seventeenth and eighteenth century travelers, describing Britain in a period for which few statistical surveys exist. Figure 1 shows an example of a statistical map created by VoB.

![Figure 1: Census Unemployment (%) in 1931, by Local Government District (Data from 1931 Census of England and Wales; boundaries mapped by the Great Britain Historical GIS).]
Secondly, we have been working since 1988, and our funding history is long and complex. However, *Vision of Britain* was created by our largest single grant, of approximately $US 1m. which came not from a research organization but from the UK National Lottery; and its purpose was to create a free resource for the general public, especially those researching local and family history. It was essential that our system be easy to use, and reach a wide audience.

Unsurprisingly, most GIS specialists focus on presenting information as maps, but this means emphasizing cross-sections, not change over time. Any truly historical resource should contain information for multiple dates, but it is surprising how few historical GIS systems can present that information as time series. Figures 2 and 3 are two such visualizations from *VoB*, and a major goal of the underlying architecture was to enable us to present statistics from sources like the census in both cross-sections and time series, without holding multiple copies of the same number.

Unemployment in Jarrow, County Durham, compared with Great Britain totals, 1927-39, computed from the Ministry of Labour Local Unemployment Index. Jarrow was famously a town of very high unemployment in the 1930s.
Our funding has required us both to hold a very wide range of content and serve several different audiences and use cases with a single computer system, and the resulting architecture is complex. Figure 4 is intended mainly to introduce three different ways of conceptualizing the system, expanded on in the next three sections, but it lists the current contents of three of the main tables in our object-relational database.

These three levels hold data values, meaning the individual statistics which are our most numerous kind of content; units, meaning formally-defined geographical sub-divisions, often with an administrative or political role but sometimes purely statistical; and places, meaning the settlements and localities where people live, and which they write about. The most detailed published account of our system is a series of three articles in the journal *Historical Methods*, and they are divided in the same way: Southall (2011) focuses on statistical data handling, ignoring most geographical issues; Southall (2012) describes our Administrative Unit Ontology, the geographical framework for the statistics; and Southall (2014a) concludes by explaining how more qualitative content has been added via “places”.

What follows can be only a brief summary but much more detail is in these and other papers. Gregory and Southall (1998) describes an earlier and very different architecture for the system. Southall (2006) provides a practical guide to using the system for local history research, while Southall (2008) focuses on our statistical visualization software. Lastly, Southall (2014b) describes specific use cases for the system: as an archival name authority; as a resource for land and watershed management; in medical research; and, commercially, as a tool to assist in establishing legal liabilities.

**Statistical Data**

The statistical content of the GB Historical GIS is very diverse:

- Statistics from every British Census of Population from 1801 to 2011, covering age structures, occupations and industries, housing conditions and so on.
- Records of births, marriages and deaths, including a large body of data on causes of death in each district of England and Wales in each ten year period between 1851 and 1910.
- Data from the annual Censuses of Agriculture since 1869, listing numbers of different kinds of animal and areas of different crops.
- The number of votes for each candidate in each electoral district in Great Britain in every parliamentary election since 1833.
- Data on economic distress, including statistics from the 20th century National Insurance system, and earlier statistics from trade union-run insurance schemes and the Poor Law.

However, unlike all other statistical databases we know of, these data are all held in just one column of a single table. This is what enables us to present the same data values in reconstructions of the original source tables, in maps and in time series; but it requires us to hold everything we know about each data value in other columns of the same table. Figure 5 provides an overview of how this works.

![Figure 5: GBH GIS Data Value Architecture](image)

Broadly, we record four and sometimes five characteristics, or dimensions, for each data value, and in most cases what is held in the data table is an identifier given meaning by a major sub-system:

- **Where**: We assume that any data value must have been gathered for a defined and bounded “unit”, rather than a vaguely defined “place”, so what we hold in the data table is an ID number for a unit defined in our Administrative Unit Ontology (AUO), as discussed in the next section. Note that the system does not require that we know the actual location of all historical units.
- **What**: Like the US National Historical GIS, we use the system of standards developed by the Data Documentation Initiative to record what is being measured. Very briefly indeed, we locate each data value within an n-dimensional hypercube, or nCube, each dimension being a Variable, and each Variable made up of Categories. For example, in the decennial cause of death information noted above, there are usually three variables: gender, age group and cause of death.
- **When**: Recording dates does not require a subsystem, but we are able to hold data for either precise dates or just calendar years, and to hold a beginning and end date when the data cover a period, such as ten years.
- **Source**: The system can hold a simple attribution for each data value, based on the Dublin Core standard, but where the data come from a British census we firstly
identify the specific census table, based on our having a complete list of all tables published by the British census between 1801 and 1961. Secondly, we can record the precise row and column a particular data value comes from, enabling us to create an online reconstruction of the tables.

- **Thanks:** A final sub-system enables us to acknowledge individuals who played a role in our making the data available, notably researchers who transcribed historical sources.

**Units**

Our approach to documenting historical geographies is similarly generalized: all “units” are defined in a single table, but a complex typology enables us to distinguish, currently, about two hundred different kinds of unit, and to add new kinds of unit whenever we want, without changing the database structure.

![Figure 6: The many different definitions of “Cambridgeshire”](image)

This permissive structure was made necessary by the extreme complexity of Britain’s changing administrative geography. Figure 6 illustrates one of the problems. Many people outside Britain will have heard of Cambridge, because of its famous university, and many will guess that Cambridgeshire is a county surrounding Cambridge. However, it is not just that the boundaries of the county have changed several times over the last 200 years; there are or were several different kinds of county, some of which existed at the same time. For example, the 1911 Census contains tables both for the Registration County of Cambridgeshire, with 215,109 population, and the Administrative County with 128,322.
Another problem is that we do not necessarily know where units were. Sometimes this is because the location of an abandoned village, and the associated parish, is now forgotten. More commonly, the sources exist but we have not yet been able to create digital boundaries for them, which we have learnt painfully is an expensive and very time-consuming task.

Figure 7 summarizes the structure of our AUO. A full entity-relationship diagram for this part of our database is included in Southall (2012), showing all the smaller tables which control the kinds of information which can be included in the main tables shown here. The system is now implemented entirely within the Postgres open source relational database, but making much use of the PostGIS object-relational extension.

Firstly, all units are defined by an entry in the Units table, but the data held there is deliberately kept very limited: a unit type, an identifier for the source the information comes from, and dates of creation and abolition, if these are known. Our most important source is Youngs (1979 and 1991).

Most information about units is held in child tables, linked by unit ID numbers, which means each unit can have any number of entries, enabling us to record change over time. The Names table holds any number of names, abbreviations and code numbers used to identify the district, recording the language and “status” for each: every unit is required to have one and only one Preferred name in each language, but we can also hold earlier Official names, with the dates of the legal change, and any number of Alternate names.

The Status table holds the finer details of what kind of unit it is. For example, Figure 1 showed unemployment by “Local Government District”, which is one of our types, but the map actually covers several different kinds of district, mainly Rural Districts, Urban Districts, Municipal Boroughs and County Boroughs. Most small towns were Urban Districts and were usually surrounded by a Rural District identically named after the town; while as a town grew it could expect to be promoted from an Urban District to a Municipal Borough, and maybe later to a County Borough. All of this is recorded, with dates of changes, in the Status table.

The Relationships table holds the one kind of positioning every unit is required to have, their location in a hierarchy being recorded through chains of IsPartOf relationships. Our commonest units are parishes, originally defined as religious units each centered on a village church but later taking on diverse administrative roles. Many of the c. 15,000 parishes in England and Wales have existed for over 1,000 years, and in that time have been parts of
many different kinds of district, all of which we try to record here, mostly based on similar listings by Youngs.

Finally, the *Footprints* table is where we hold boundary polygons: each polygon can have a start and end date, enabling us to record sequences of boundary changes. However, because of the cost and difficulty of reconstructing historical boundaries, only about half our units have polygons. Most of the rest have a simple point coordinate, held in the Units table.

**Places**

The above structures enable us to hold many millions of statistical data values in a single table, and to present them flexibly in maps, graphs and table reconstructions. However, and especially under our National Lottery funding, we wanted to hold a wider range of content:

- **Scanned images of historical maps.** See Southall (2014a) and Southall and Pridal (2012) for our work with these. All maps are georeferenced, enabling us to link from places to maps which cover them.

- **Descriptive Gazetteers.** Especially during the nineteenth century, publishers created lengthy gazetteers describing named places, sometimes in a single sentence but sometimes in “entries” which are 100,000 words long. We have computerized, or obtained through collaborators, four major gazetteers totaling over 90,000 entries and around seven million words.

- **Travel writers.** *Vision of Britain* is now the largest online collection anywhere of historical British travel writing, featuring 28 narratives of journeys around Britain, mostly book length. These include William Camden’s *Britannia* (1607), the first county-by-county survey of Britain, by itself over half a million words; Celia Fiennes’ *Through England on a Side Saddle*, describing remarkable journeys by an unmarried woman in the 1690s; and Daniel Defoe’s *A tour thro’ the whole island of Great Britain*, originally published between 1724 and 1727 and describing Britain on the verge of the Industrial Revolution. Geographical names are marked up following the principles of the Text Encoding Initiative (Sperberg-McQueen and Burnard, 2002).

![Image](image.png)

*Figure 8: Places mentioned by Celia Fiennes when describing a journey from Coventry to London in 1607*

The marked-up texts are then analyzed by a program which builds a separate concordance table with one row for each identified place-name instance, storing the name itself, the text it
appears within, the position within the text, and a place identifier which we can link to a geographical coordinate. This enables four kinds of functionality in the web site. Firstly, the online text contains embedded hyperlinks, so users can click on the name to find out more about the place. Secondly, from the concordance table we can create simple maps, such as Figure 8, showing the places mentioned in a particular piece of text. Thirdly, we can link from our page for a particular place to all the travelers who wrote about it. Finally, the particular name forms used by travelers become additional variant names for the places in our gazetteer.

![Diagram](http://www.VisionOfBritain.org.uk)

Figure 9: Integrating qualitative information at the “Place” level

However, it was obvious we could not create these links to units in the AUO: a typical town will have been the headquarters of ten to fifteen different units over time, and it would be a nonsense to link a statement like “I now arrived in Portsmouth” to any one of them. We therefore constructed a separate gazetteer of “places”, starting by including every settlement which gave its name to an Urban District or Parish, but later adding additional locations mentioned by travel writers, or simply having an entry in one of the descriptive gazetteers. This then became the most important geographical backbone for the whole system.

Figure 9 shows how we then harvested place names from all our textual sources to add more variant names; and note that this is the same table of Names already described for the AUO, so all the names listed in census reports end up also included here. We are starting to add place names appearing on old maps, but of course that harvesting cannot be automated. The end result is a compact gazetteer of Britain’s towns and main villages, including an exceptionally rich set of variant names each linked to the historical source in which it appears. For example, both Google and Bing autocorrect “Norwich” to Norwich, but VoB tells us this name was used by Celia Fiennes for both Norwich and Northwich in Cheshire.

User Experience

So far, we have described how we have assembled and intricately interconnected a mass of different kinds of historical source material, but how does our web site look to a user? One answer, of course, is to see for yourself:

http://www.VisionOfBritain.org.uk
Data on site usage from Google Analytics shows that most users enter the site at a geographically-specific page, most commonly a Place Page such as Figure 10. This immediately demonstrates how we have integrated different kinds of content: on the page, an excerpt from an old map and a descriptive gazetteer; but with links to travel writers, statistical data and, simply by clicking on the map excerpt, our historical map collection.

Figure 10: The place page for Portsmouth within A Vision of Britain through Time

Figure 11 demonstrates a key feature of our site’s performance: it is designed to be very easily found via search engines. Obviously, we do not appear very high in the list of search
results for “London”, but for villages and small towns we are often in the top half of the first page of results. This is a direct and planned consequence of our geo-semantic architecture. The “spiders”, such as Googlebots, which search engines use to index the web, work only with text, not graphics, and can only follow links, not complete forms. GIS-driven maps embedded in web pages are, technically, just large web forms consisting mainly of graphics, so most GIS-based web sites are almost invisible to search engines. The IsPartOf relationships in our AUO create a massively interconnected web site, which Googlebots love.

![Bar chart showing the number of parishes](image1.png)

Figure 11: Source of first-ranked results from searching google.co.uk for "history of <name>" for all Herefordshire parishes

Finally, figure 12 shows how we achieved the goals set for us by the National Lottery: after some years of usage slowly building up, we now receive around one and a half million visitors each year, far more than most academic historical sites. Advertising income from the site has become a significant source of income for our team.

![Line chart showing monthly unique visitors](image2.png)

Figure 12: Monthly “Unique Visitors” to A Vision of Britain through Time, 2004-2016
Conclusion

Our experience suggests the following more general conclusions for creators of digital historical atlases:

Firstly, know your audience and establish a dialogue with them. Our funding from the UK national lottery required us to be interesting to a broad public audience, not just other university historical researchers. Establishing dialogue is far easier online than in traditional publishing, although with a popular site you may receive excessive amounts of automated “spam” email. The lottery funded some formal usability testing, done for us by specialists, but do-it-yourself usability testing is far better than none (Krug, 2009). Diagnostic tools such as Google Analytics help you observe exactly how users move around your web site. It is worrying that most academic web sites gather no usage data at all.

Secondly, do not confuse a printing press with a library: an online GIS is not an atlas. It is easy to find examples of historical web sites built directly using software like ArcGIS, often built by projects led by archaeologists, but even if the software is usable without formal GIS training, users without a conceptual understanding of geographical information and cartography will all too easily create nonsense maps, such as choropleth maps of absolute values. Our task as atlas authors is to enable our users to be readers, to whom we present maps and other materials which are interpretations. It is more important to enable users to drill down to the underlying data than to let them create their own alternative maps. To achieve our goals, we make extensive use of ArcGIS for offline editing of digital boundaries, but the online system combines open source components including Postgres, MapServer, GeoWebCache and OpenLayers.

Thirdly, despite the Oxford English Dictionary, atlases are not just “books of maps or charts”: I contributed to the Atlas of Industrializing Britain (Langton and Morris, 1986) and co-edited An Atlas of Industrial Protest in Britain (Charlesworth et al, 1996), and both contain as many pages of text as of maps, and very many graphs, often time series. The remarkable three-volume Historical Atlas of Canada (e.g. Holdsworth and Kerr, 1990), with a larger budget, has less text but most pages are rich assemblies of information with much besides maps. Building an atlas around a computerized database means that the information assembled can and should be presented in many different ways. A fundamental difference between our project and the US National Historical GIS, both focused on historical censuses, is that their Social Explorer web site (http://www.socialexplorer.com/) keeps censuses separate and so visualizes data only through maps, while Vision of Britain can present each individual data value, stored once, as part of a map, or a time series, or within a reconstruction of the printed table it came from, or within a downloadable dataset. That said, limited funding has restricted the range of visualizations we offer: we hope to explore animation, and the diverse approaches to graphically representing change through time assembled in Rosenberg’s Cartographies of Time (2012).

Fourthly, in a large system such as ours, creating non-spatial frameworks for information is as important as, and maybe harder than, creating spatial frameworks; for example, our data and source documentation sub-systems, the former being essentially a statistical domain ontology. Wherever possible atlas projects should base their work on existing standard works, as with our use of Youngs (1970 and 1991), and more generally we need to understand and follow the principles of information science, or library science (Morville and Rosenfeld, 2006), not just geographical information science. A major online historical atlas will need to draw on many more disciplines than just history, geography and cartography.
Fifthly, few people sit down and read an atlas cover-to-cover. They are primarily works of reference, and today most users seeking reference material use a search engine such as Google. Few academic projects seem to have given any thought to how far the information within their site can be reached by Google’s index-building “bots”, and the graphical nature of cartography and GIS poses special problems. However, we showed above that this was a solvable problem, albeit requiring a hybrid geospatial/geo-semantic architecture. One reason we did well was that our lottery funding required a focus on “accessibility” by those with disabilities, and especially the blind and partially-sighted. VoB was therefore designed to work well not just with conventional web browsers but also with screen readers using text-to-speech systems such as Microsoft Narrator or Apple’s VoiceOver. These ignore graphics, and generally experience the web in a very similar way to Google’s bots (Walter, 2008), so a site which works well for the blind works well with search engines.

Building a large spatially-enabled online historical resource, such as A Vision of Britain through Time, poses many challenges and is certainly more expensive to author than a traditional paper atlas. However, traditional atlases are then much more expensive to print, especially in color, so few non-specialists will be able to buy them. This chapter has described some of the challenges, and the solutions we found, but it has also shown that an online historical project can reach a far larger and wider audience than any paper atlas, serving the millions of people wanting to know more about how their home towns and villages, and their ancestors’ birthplaces, have changed through the centuries.

References
Camden, W. (1607) Britain, or, a Chorographicall Description of the most flourishing Kingdomes, England, Scotland, and Ireland. Our online version is based on Dana Sutton’s hypertext critical edition of the English translation by Philemon Holland of Camden’s Latin text, online at http://www.philological.bham.ac.uk/cambrit.


