19 Spatial Frameworks for Historical Censuses

The Great Britain Historical GIS

Ian Gregory
Humphrey Southall

Introduction

Although the first censuses were the only practical means of surveying a nation's population, the development of sample survey methodologies means censuses are increasingly justified by the specifically geographical detail they uniquely provide. The same is arguably true of historical census studies: a two per cent sample taken from many millions is quite sufficient to study national class or household structure. Once sample sizes reach ten per cent, local detail starts to emerge, and arguably the only sensible reason to work with a complete census is to study the fine geographical structure of past societies.

Whether we look at the published reports or the individual-level returns, much of the information in every census is geographical: tables of national totals are a mere preface to endless tables at state, province, county, district,

Humphrey Southhall is Programme Director of the Great Britain Historical GIS Project (GBHGIS) and a Reader in Geography at the University of Portsmouth. Ian Gregory is a Senior Fellow at Portsmouth, working on the GBHGIS.

For more information on the Great Britain Historical GIS project, see the GBHGIS web site at: http://www.geog.port.ac.uk/gbhgis/.

parish, commune or township level. Census schedules identify households through a similar spatial hierarchy down to the street and building; individual subjects of a census are usually doubly located, by residence and birthplace. One large caveat remains: the process of anonymising microdata often removes not just names but most locational detail, destroying geography. However, this is less true for historical microdata.

Unfortunately, most of the geographical detail in historical censuses is uninterpreted, consisting of place-names only meaningful to local experts and defunct administrative areas. We need a map, and today this usually means a computerised mapping system, or a Geographical Information System (GIS). Preparations for modern censuses normally include the construction of extremely detailed mapping systems, but census geographies have changed drastically over time and historical researchers must generally develop their own. Mappers of historical census data at the provincial or U.S. state-level can often use modern base maps, but reconstructing the more detailed geography of past censuses involves research projects as large and long-term as any project more directly concerned with microdata.

This chapter describes the Great Britain Historical GIS Project, which developed in collaboration with the research into the 1881 British census described elsewhere in this volume. Such a project could not be justified simply to map a single source, even one as large as the 1881 microdata. We had to offer a more general solution to the mapping needs of British historical demographers, and we had a research agenda of our own—to identify long-run trends from macrodata—meaning tables from the published census reports since 1801. These other goals necessitated constructing a detailed record of changing geographies. Simply constructing a detailed base map for 1881 was only possible by first computerising good quality maps published in the late 1900s and then carefully interpreting and "back-projecting" the length lists of inter-censal boundary changes that appeared in the reports of the 1911, 1901 and 1891 censuses. It was over this period that the modern system of local government in Britain was created through the massive rationalisation of essentially pre-industrial ecclesiastical units. The resulting Great Britain Historical GIS (GBHGIS) is the largest time-variant GIS yet built, tracking the evolving boundaries of over 15,000 parishes, as compared to only 3,500 U.S. counties. It adopts a particularly elegant solution to representing changing boundaries, has a close relationship with an equally large microdata project, and the resulting Digital Boundary Data (DBDs) are available on-line through a specialised dissemination facility. All these features justify its use as an exemplar, but the chapter concludes with a survey of historical boundary mapping projects in Europe and North America. There is no necessary connection between these projects and historical census research, but most originated in the needs of historical demographers.

Origins of the Great Britain Historical GIS

Many have attempted to define GIS. Perhaps the most appropriate definition here is Goodchild's: A GIS is best described as a system which uses a spatial database to provide answers to queries of a geographical nature. The generic GIS can thus be viewed as a number of specialised spatial routines laid over a standard relational database management system.[1] A GIS is therefore a database management system with an added spatial component, linking two types of information: attribute data which are the statistical data, and spatial data providing locations, usually in the form of points, lines and areas.[2] A historical GIS adds temporal data.

The GBHGIS is therefore not simply a mapping system. The spatial data consist primarily of boundary lines which can

---

[1] Goodchild M.F. 1985. "Geographical Information Systems in Undergraduate Geography: A Contemporary Dilemma." *The Operational Geographer* 8, pp. 34-8 (35).

[2] Cowen, D., "GIS versus CAD versus DBMS: What are the Differences?" in D. Peuquet and D. Marble (eds.) 1990. *Introductory Readings in Geographical Information Systems*, pp. 52-62.

be assembled into polygons describing administrative areas, while the temporal data are precisely-dated boundary changes. However, these data exist mainly to give meaning to our attribute data, for which the censuses, including 1881, are the single most important source. These attribute data are held in a closely linked relational database, the Great Britain Historical Database (GBHDB).[3] Our aim has been to create a common framework within which spatially referenced official statistics published since the early nineteenth century can be integrated. This will enable us not simply to map the data, but to develop new and powerful ways of manipulating them to answer previously unanswerable questions, particularly relating to long-run trends.

Our parish-level historical GIS has developed since 1996 in parallel with the 1881 census research, but Humphrey Southall had been assembling statistical data concerned with the origins of Britain's "north-south" divide since the 1970s. In 1989-91, outside funding permitted construction of a systematic database of pre-1914 unemployment data, poor law statistics, small debt data and other indicators of hard times. Like census reports, all this information consisted of numeric tables whose left-hand column contained place-names.[4]

Interpreting the resulting database was problematic because so many concerned geographical entities that either no longer existed or had changed their boundaries significantly. Our immediate solution was to build gazetteers enabling us to alculate county and regional totals, but fully exploiting the wealth of spatial detail clearly required mapping. Our initial focus was not parishes but the Poor Law Unions of England and

---

[3] British law includes the offence of 'GBH' or gross bodily harm.

[4] Gilbert, D.M., and Southall, H.R. 1996. "Indicators of regional economic disparity: the geography of economic distress in Britain before 1914" in C. Harvey and J. Press (eds.) *Database Systems and Historical Research* (Basingstoke), pp. 140-6. Southall, H.R., and Gilbert, D.M. 1996. "A Good Time to Wed: marriage and economic distress in England and Wales, 1839-1914," *Economic History Review*, XLIX, pp. 36-58.

Wales, designed to represent a market town and its hinterland.[5] For this initial modest GIS we developed methods for systematically representing changing boundaries within a computer, allowing us to link a dataset from any date to an accurate representation of the administrative units for that date. This powerful system attracted many collaborators, most notably the 1881 project who supplied additional data and helped fund us, but wanted a record of parish boundaries; about 15,000 units as compared to 630 Unions.

We also obtained substantial direct funding for the extensive digitisation of published statistics from the census and vital registration, as well as GIS construction; this the work of a second team at the Queen's University, Belfast, using optical character recognition technology. The current attribute database includes:

- **Labour market statistics:** trade union unemployment 1851-1912, and government data 1902-39; poor law statistics 1859-1939; small debt plaints, mainly 1868-1913; urban wages 1845-1906.
- **Population census** (usually at sub-county level): age structure, 1851-1961; occupational data 1841-1951; county of birth 1851-1951; parish-level populations for most censuses 1801- 1951.
- **Mortality** (in progress): decennial age and cause of death, by decade, 1851-1930; annual mortality counts 1851-73; quarterly sub-district counts for census years 1871-1911.[6]

We have access to British electoral statistics since 1885 and sample data from collaborators for earlier surveys of Britain back to the Domesday Book of 1086. Detailed documentation is at: http://www.geog.port.ac.uk/gbhgis.

While this book focuses on microdata, we suggest that much can still be done by re-working published aggregates. In

[5] Cooke, P. (ed.) 1989. *Localities: the changing face of urban Britain*.
[6] Contributors to our attribute DB include Danny Dorling, David Gatley and Robert Woods. Other donations welcome!

particular, our methodologies enable redistricting of data published for many different systems of areas to a single constant geography, revealing long-term trends. There is also great potential for linking microdata and macrodata. Bizarrely, the only legal access that exists to individual-level mortality data

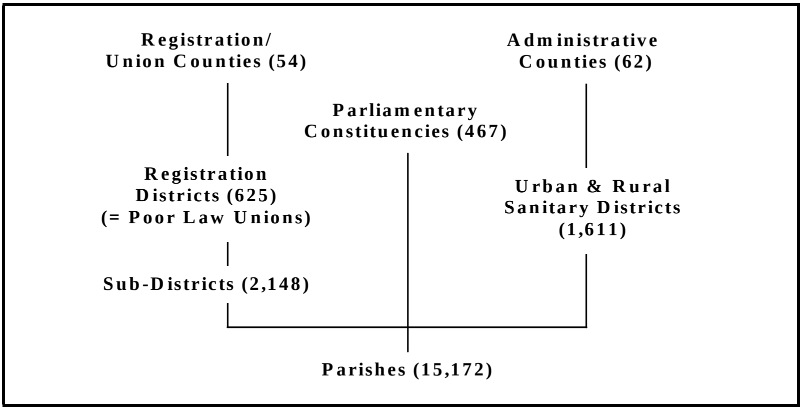


Figure 19-1. Principal Statistical Reporting Units in England and Wales, 1881

(Constituencies 1885; Local Government/Sanitary Districts 1891)

for Britain is through buying copies of individual certificates. Historical research must therefore be based on aggregate data, though the previously un-exploited sub-District data cover 2,175 areas. The only published data for these cover age-structure, permitting the calculation of Standardised Mortality Ratios, so we are using 1881 micro-data to compute occupational statistics, relating mortality to locality in unparalleled detail.

Building the GIS

Information in our statistical database relates to geographical entities, mostly formally defined areas. The 1871 census reported information for fifteen distinct geographical hierarchies but most had specialised functions.[7] The most important geographies are shown in Figure 19-1:

> **Poor Law Unions/Registration Districts (PLUs/RDs)** (and sub-Districts): Unions were created in 1834, dividing England and Wales into about 630 unions combining towns and hinterlands. Civil registration, introduced in 1837, used the same units as Registration Districts, and sub-divided them into sub-Districts. Abolished 1930.
>
> **Sanitary/Local Government Districts (LGDs)**: From the 1870s, approximately 1,700 LGDs were formed. In general, urban parishes became County or Municipal Boroughs, or Urban Districts, the remaining rural parishes in each Union becoming a Rural District. Abolished 1974.

All geographies used parishes as building blocks, except in towns where a single parish might be sub-divided for vital registration and elections. Parishes were ancient ecclesiastical units, but the modern Parish system was created only in the 1860s (after the first censuses) and then extensively modernised by the Divided Parishes Acts of 1876 and 1882.[8] However, the largest single problem is not apparent from examining the names of districts in statistical tables, or even a single set of maps: the boundaries of all these units were constantly revised. Census and vital registration reports list boundary changes back to 1876 for parishes, and 1861 for Registration Districts, from which we have built a database listing 383 changes to Unions/Registration Districts, 5,874 changes to Local Government Districts and 26,944 changes to parishes. Unfortunately, working out the consequences of these changes, especially undoing the late nineteenth century rationalisations to

---

[7] *1871 Census Report*, Appendix C: Territorial Sub-Divisions of England, p.175.

[8] Lipman, V. 1949. *Local Government Areas 1834-1945* (Oxford); Hasluck, E. (1936) *Local Government in England* (Cambridge).

create an accurate 1881 map, meant establishing the effect of each of many thousands of changes on actual boundary lines. Creating a systematic record of these changes meant building a true temporal GIS.

Geographical Information Systems technology has developed to meet the needs of marketeers and utility companies, and available commercial software makes little provision for incorporating a time dimension.[9] While other projects have constructed temporal GISs from scratch, we needed not just to map change but link to attribute data and perform analysis. The GBHGIS is therefore based on existing software, the well-known ArcInfo package, using bespoke software written in Arc Marco Language (AML) to manage changes over time.

Like most GIS software packages, ArcInfo stores its data in distinct coverages or layers, each usually consisting of some combination of points, lines (arcs) and areas (polygons). All can have linked attributes, those for polygons being stored in label points.[10] This feature allows us to store date-stamped information about a unit separately from temporal data about its boundaries, and therefore store the complete history of an administrative geography in one single master coverage. Taking transfers as an example, two arcs (or sets of arcs) are used in the master coverage, one representing the pre-change boundary and one the post-change boundary. Figure 19-2 gives the example of a central area being transferred from “Anarea” to “Elsewhere”

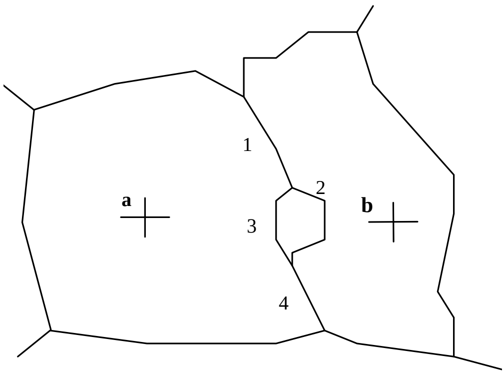
---

[9] Temporal GIS remains under-developed. See Egenhofer, M., and Golledge, R., (eds.) (1998) *Spatial and Temporal Reasoning in Geographic Information Systems* (Oxford); Wachowicz, M. 1999. *Object-Orientated Design for Temporal GIS*; Vrana, R. 1990. “Historical Data as an Explicit Component of Land Information Systems” in D. Peuquet and D. Marble (eds.) *Introductory Readings in Geographic Information Systems*, pp. 286-302.

[10] Peuquet, D., and Marble, D. 1990. “ArcInfo: An example of a contemporary Geographic Information System” in D. Peuquet and D. Marble (eds.) *Introductory Readings in Geographic Information Systems*, pp. 90-9.

on the September 1, 1894.

Figure 19-2. Date-Stamps Used to Represent Transfers



Label Point Name and Date Stamps:

	Name	Start Date	End Date
a	Anarea	0/0/0000	0/0/5000
b	Elsewhere	0/0/0000	0/0/5000

Arc Date Stamps:

	Start Date	End Date
1	0/0/0000	0/0/5000
2	0/0/0000	1/9/1894
3	1/9/1894	0/0/5000
4	0/0/0000	0/0/5000

Label points are represented by crosses and identified by letters, while arcs are identified by numbers. Features in existence when a unit was formed are date-amped 0/0/0000, while those in existence when the unit was abolished are date-stamped 0/0/5000.

Building this system involved extensive use of primary sources. Our foundations were the first unproblematic maps of administrative areas, published in the late 1900s at 1:126,720 scale. These were out-of-copyright, used a single map projection for the whole country and gave a sensible compromise between our needs for detail and rapid completion.[11] Figure 3 shows part of Wales, including boundaries for parishes, Poor Law Unions, rural and urbandistricts, and Newport County Borough. Once these maps were digitised, joined together and checked against the 1911 census report, we had to add boundary changes. Except wheretransfers were of hole parishes, this meant finding maps. This was problematic for before 1900: the national mapping agency began mapping boundaries in the 1840s and finished only in 1893, after a period of massive change.[12] Our system relies on unpublished maps prepared by the census office in the 1860s and 1880s, supplemented from the earlier tithe maps (Public Record Office classes RG.18 and IR.105; one team member spent two years there).[13]

Digitising the 1900s base maps proved a small part of the project; most of the three and a half years construction period being the intricate assembly of a record of change. Unfortunately, time-consuming manual checking was needed to ensure that whenever one boundary segment disappeared it was replaced by another to maintain the integrity of all resulting polygons/areas. So far, the boundary change GIS is limited to England and Wales—with a separate static boundary map created from *Black's Atlas* to plot Scottish 1881 data—but the main system will be extended to Scotland.

Figure 19-3 shows one end product—our 1881 Hertfordshire parishes linked to 1881 microdata. Without the microdata we could not measure movements out of Hitchin, a small town at top centre, but without the GIS we could not show destinations. The static 1881 map is just one output from our temporal GIS. We are beginning our own programme of

---

[11] Oliver, R. 1993. *Ordnance Survey Maps: A Concise Guide for Historians.*

[12] Owen, T., and Pilbeam, E. 1992. *Ordnance Survey: Map Makers to Britain since 1791.*

[13] Kain, R., and Oliver, R. 1995. *The Tithe Maps of England and Wales* (Cambridge).

analytic research, initially focused on mortality decline. Past UK research terminates in 1911 when reporting switched from Registration to Local Government Districts, but the GIS can bridge that gap. Another copy of the system has been supplied to UKBORDERS, a specialised service established to disseminate modern census DBDs. Their software already permitted users to specify part of the country, type of unit and GIS file format; it has been extended so that users can also select precise dates: http://edina.ed.ac.uk/ukborders.

Figure 19-3. People Born in Hitchin, by Parish: Hertfordshire, 1881
(Supplied by Mark Allen, 1881 project)

National Historic Boundary Mapping Projects–A Survey

The GBHGIS is the largest historical GIS but scarcely the first. Many demographic projects have constructed simple static digital boundary maps, either for a single census or in the optimistic belief that boundaries never changed; we know of at least three digital maps of British Registration Districts. The following survey emphasises more sophisticated approaches, and draws on a recent workshop on historic boundary mapping involving seventeen European countries plus the US.[14]

The earliest attempts were both American, but neither quite worked. The Atlas of Historical County Boundaries project at Chicago's Newberry Library began in the late 1970s using automated mapping software developed by the University of Wisconsin, Madison. However, technical difficulties led to the atlas project reverting to traditional methods, and they are only now starting to computerise. However, their boundary research is arguably the most detailed anywhere, using textual descriptions of boundaries from court records, some pre-1776, which must be painstakingly mapped. Each state volume provides both maps for each county showing all boundary changes, and a state map for each census date.[15]

The Great American History Machine, originally developed at Carnegie Mellon, again pre-dated modern GIS technology. It links county-level statistics from every U.S. census—plus all presidential elections—to a county base map which is dynamic in that it allows for the progressive subdivision of the country as the frontier moved west. The Newberry research shows the process was more complex. The original system ran only on Unix workstations. More recently a Windows version was developed at the University of Maryland but again not released commercially.[16]

---

[14] "Mapping Europe's Historic Boundaries and Borders," supported by the European Science Foundation. Extensive surveys of projects and archival resources are at http://www.geog.port.ac.uk/hist-bound.

[15] Basoglu, U. and Morrison, J. 1978. "The Efficient Hierarchical Data Structure for the U.S. Historical Boundary File" in Dutton, G. (ed.) *Harvard Papers on GIS: Volume 4* (Reading, Mass.); For current information see http://www.newberry.org/nl/scholl/L3ratlas.html.

[16] Miller, D., and Modell, J. 1998 "Teaching United States History with the Great American History Machine "*Historical Methods*, 21, pp. 121-134.

The Canadian Families Project, primarily a microdata project, have mapped census tracts, the main technical feature being that the census geography was unrelated to administrative units, necessitating much primary research.

In Europe, vast changes in national boundaries plus wartime destruction of archives complicate such research. Most progress has been made in Scandinavia and the low countries. The NLKAART system maps boundary changes for Dutch municipalities: there were 1,227 municipalities in 1830 but only 600 in 1994. NLKAART is written as a SAS/GRAPH application whose data consist of two types of files: concordances store information about the names and dates of each place, with a new entry every time its co-ordinates change. The co-ordinate files stores municipality boundaries between each change. These files can be linked for any place, and data added to create shaded maps. The data have recently been converted to work with a modern GIS package, MapInfo, by simply extracting every possible arrangement of units and storing these as separate MapInfo files; the total storage requirement was still under 100Mb.[17]

The Quantitative Database of Belgian Municipalities covers evolving municipalities and villes from 1796 to 1970; in 1830, there were 2,498 municipalities and 89 villes. This system links Atlas GIS software to a large database stored in Lotus spreadsheets, covering each province via a sequence of static maps, and using a concordance to build a map of Belgium for any given date using the most appropriate set of provincial maps; there are periods of rapid change when this breaks down.[18]

---

[17] Boonstra O (1994) "Mapping the Netherlands, 1830-1994: The Use of NLKAART" in Goerke, M. (ed.) *Coordinates for Historical Maps* (Gottingen), pp. 156-161.

[18] Vanhaute, E. 1994. "The Quantitative Database of Belgian Municipalities (19th/20th centuries): From Diachronic Worksheets to Historical Maps" in Goerke M (ed.) *Coordinates for Historical Maps* (Gottingen), pp. 162-175.

In Sweden, changing historic boundaries for c. 2,500 parishes were mapped in the early 1990s by the National Archives as a component of their NAD meta-data system, using bespoke software, and the results used extensively with the Umea Demographic Database. Norwegian municipality boundaries (450-725 areas) were similarly mapped back to 1837 in the 1970s by the Norwegian Data Centre for the Social Sciences.[19] Danish historic boundaries have been researched but not computerised.

Germany is once again the largest European country, by area and population, but its size, complex history and federal structure have inhibited large scale mapping projects. The Deutsche Grundkarte project aims to collaboratively map the whole country, using similar methodology to the Belgian system, but progress has been difficult.[20] A separate historical atlas project for Prussia simply holds snapshots for periods of administrative stability.[21]

Elsewhere, in eastern and southern Europe, there are many interested researchers but little completed boundary mapping. One exception is Portugal, where Luís Silveira has reconstructed boundaries for reconstructed boundaries down to parishes for 1801, 1826, 1834, 1842 and 1855. Since then there has been little change. The EUROGRAPH project at the Teleki László Institute, Budapest, is mapping historic boundaries, mostly at the district level but sometimes for communes, not only for Hungary but also for several adjacent regions. For example, the project has created a detailed map of Transylvania in 1910.

---

[19]Oldervoll, J. 1994. "Mapping – in a Norwegian Perspective" in Goerke, M. (ed.) *Coordinates for Historical Maps* (Gottingen), pp. 184-9.

[20] Ott, T. and Swiaczny. 2000, in press. "The Analysis of Cultural Landscape Change: A GIS Approach for Handling Spatio-Temporal Data." *History and Computing.*

[21] Winnige, N. 1999. "*Alphabetisierung in der Frühen Neuzeit, oder: Wie visualisiere ich raumbezogene historische Daten*?", in Ebeling D (ed.): *Historisch-thematische Kartographie. Konzepte, Methoden, Anwendungen* (Bielefeld), pp. 82-97.

Conclusion

Systematic spatial frameworks are integral to modern census taking, and are equally essential to effective historical research using complete censuses or large samples: the GBHGIS was the largest piece of infrastructure created to analyse the British 1881 census. Equivalent sources and relevant preliminary research exists for most other parts of Europe and North America. However, such projects face two major problems. First, available commercial software is ill-suited to temporal GIS. Second, and partly in consequence, historical GIS construction is very expensive: the GBHGIS has cost over $1m., albeit including much digitisation of published statistics.

Until software improves, we should perhaps concentrate on assembling place-name gazetteers and thesauri, recording hierarchic relationships between units plus simple locations of the lowest-level units; for example, locating parishes via co-ordinates for churches. From such co-ordinates, GIS software can generate surrounding areas crudely approximating each parish.[22] Data on hierarchies, often from census reports, then enables us to better approximate districts as assemblies of parishes. All maps are approximations, and the maps so generated will, arguably, often be "fit for purpose."

The construction costs of true temporal GIS systems may be hard to justify simply to assist demographic research, but they have other uses. They can contain both information on the history of every individual town and village, and a locational framework enabling a wide range of graphs and maps to be generated, making the past more comprehensible to school children and local history researchers. One eventual goal for the GBHGIS is to become just such a national resource for popular history.

[22] Haggett, P., et al. 1977. *Locational Models*, pp.436-9.