A Vision of Britain through Time: Making long-run statistics of inequality accessible to all

Humphrey Southall

Britain has been gathering systematic statistics about its localities for two hundred years, through the census, vital registration and other mechanisms. However, the resulting data are useless to researchers interested in really long-run trends at the local level, and simply inaccessible to the bulk of the population:

- Few public libraries – let alone homes – have complete sets of census reports back to 1801, or even copies of the Registrar General's *Statistical Reviews* for the mid-twentieth century. The academic and copyright libraries which do are closed to most people.
- When you do find the reports, you see shelf-yards of fat volumes, often disintegrating, full of figures mainly about geographical areas within Britain – but data for your area are scattered right through the volumes. Census reports sometimes provide comparative figures for the previous census, or even the last-but-one, but establishing long-run trends usually means finding and copying data from each report in turn.
- Once you figure out which volume to look at, it will consist of table after table, in each of which the left hand column – or two – contains place-names, some of which will be familiar. However, these are really the names of administrative *areas*, not of *places*, and you soon discover that these lists keep changing. In fact, the system of districts and unitary authorities used by the 2001 census was created only in 1996. Earlier comprehensive changes in reporting geographies happened in the mid-nineteenth century, when ancient hundreds, wappentakes and so on were replaced by Registration Districts; in 1911, when Registration units were replaced by Local Government Districts; in the mid-1930s, when the County Reviews consolidated local government areas; and in 1974/5, when a new system of much larger districts, plus Metropolitan Counties, was created. At this point even academics give up, before learning there was also a constant trickle of individual boundary changes.
- There are also, of course, variations over time both in what data were gathered and how they were classified. Changing classifications are particularly a problem with occupations – and therefore social class – and cause of death.

- ... and, of course, all of this stuff is basically tables of numbers, which are really boring unless you are a complete anorak.

This article is about how and why I tried to change this situation, obtaining a National Lottery grant to build a web site opening up Britain's 'statistical heritage', and presenting as clearly as possible long run comparisons between areas on topics such as unemployment, class and infant mortality. Although this was not a Radical Statistics Group project, it involves a number of Radstats members.

Committing GBH

The Great Britain Historical Geographical Information System (GBH GIS) project developed out of my personal obsession with the origins of Britain's north-south divide. This is usually traced back to the inter-war depression, and seen as the result of some kind of dis-equilibrium process, but my earliest work used records of trade union welfare benefit systems to show that unemployment during late 19th century and early 20th century recessions, such as those whose greatest impact came in 1879 and 1909, had a very similar geographical pattern to the 1930s. My paper on the origins of the depressed areas concluded by quoting the economist Pigou:

> 'A nation which concentrates its forces upon the manufacture of the instruments of industry courts thereby a relatively heavy burden of unemployment.'

I went on to say:

> 'This burden was therefore part of the price Britain paid for an apparently prosperous industrial system based on exports rather than indigenous demand, and it seems clear that the burden fell most heavily, both then and subsequently, on those regions which created her prosperity.'

My article included a lengthy review of studies, mainly textbooks, which asserted that industrialisation brought prosperity to the 'factory districts', which lasted until the First World War, and showed they presented no actual evidence. That appeared in a fairly high-profile academic journal – the *Economic History Review* (1988) – fifteen years ago, and in some obvious senses was 'well received'. However, it had little impact on what went into text books, or on the untaught history that people 'just know'. How could I change this, and get across the longevity of this geography of disadvantage that simply requires another kind of explanation?

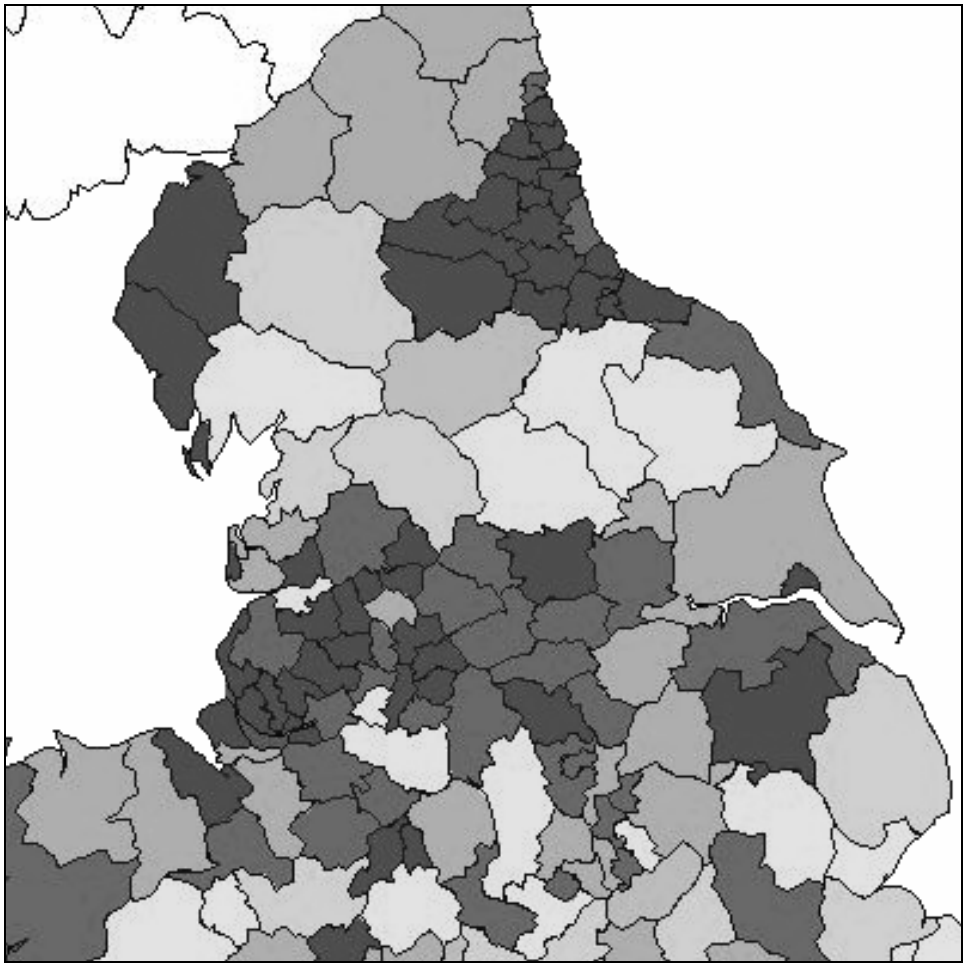
One answer was to gather more data, funded by a series of increasingly large research grants. The first, funded by the Leverhulme Trust in 1989-91, computerised a great deal more trade union statistics, and also covered three other measures of economic distress in the pre-1914 period: obviously, the Poor Law; slightly less obviously, legal statistics for 'small debts' of under £20, generally owed by working class households to shop keepers; and least obviously, the number of people getting married. Statistics of the Poor Law were dominated by payments to farm workers during winter, funded collectively by the local land owners; in the towns, it mainly provided for the sick and the elderly (the system was always more about cash payments than the workhouse, despite the many myths). The problem with small debt data is that in the industrial areas both shop-keepers and their customers understood how the trade cycle worked, so 'putting it on the slate' would be managed carefully. The system is still interesting because debts under £20, generally owed by the working class, were treated very differently from larger debts owed by the middle class: the latter could declare themselves bankrupt, but small debtors had to pay up or go to prison (Johnson, 1993).

If small debt data record mistakes by shop keepers and their working class customers who misjudged the economy, marriage data are fascinating because they show, in a sense, the extent to which working class communities had to adapt themselves to an unstable industrial system. The number of marriages is essentially a record of household formation, only possible once the couple had established a degree of economic independence, through relatively steady jobs and accumulated savings. This meant that in areas dominated by cyclical unemployment recessions led inevitably to the postponement of marriage. To take the example of Bolton in Lancashire, the total numbers of marriages doubled between 1842 and 1845 as the economy recovered, and while those where the groom was an engineer trebled. Even middle-class marriages increased by two-thirds (Southall and Gilbert, 1996). The Poor Law also repays closer study: with enough data, the impact of both highly seasonal payments in farming areas and the impact of long-run political pressure to reduce relief payments can be filtered out, and the remaining geography is dominated by the impact of cyclical recessions. It is not widely understood that in the 1920s and the 1930s 'the dole' was poor relief, most workers exhausting their entitlement under the fairly new National Insurance system: political changes meant that relief to unemployed industrial workers was now seen as legitimate, so relief levels in the industrial north rose rapidly (Southall, 1998).

The original work on trade union data used branch-level statistics, and without data on members' home addresses had to treat each branch as a point based on where they met – generally the back rooms of pubs. Marriage and poor law data, however, were clearly gathered for Registration Districts and Poor Law Unions. These were administered by two different sets of officials but had identical boundaries, dividing England and Wales up into a changing set of roughly 630 districts. Studying them was hard because we had no map of their boundaries, so I applied in 1993 to the ESRC for funding for a one-year and one-person project to build a computer mapping system to work with the existing statistical database. That project kept growing: it is still running, and at the time of writing employs 17 staff, 9 full time. We seem to have established ourselves as something called the Great Britain Historical GIS Project, but this has no legal existence and it has all been funded by a series of 24 grants from 14 funding bodies, each for a 'project' with a different name. The 'GBH' tag is a bit dubious; we started using it as an interim name when we were hoping for a grant from the Millennium Commission for a project covering the whole collection of islands, then called the Millennium Atlas Project (MAP). If we ever do cover Ireland, the overall name will still be a problem as the term 'British Isles' is unacceptable in Ireland.

The project moved on from mapping Poor Law Unions/Registration Districts to mapping the local government districts used for most statistical reporting between 1911 and 1974, but the largest task has been mapping the changing boundaries of the 15,000 or so parishes of England and Wales. About the only statistics we have for parishes in most years is the total population, but knowing these enables us to much more accurately redistrict data for higher-level units. For example, the 1931 census reported unemployment in 1,800 local government districts across England and Wales, ranging in size between the County Borough of Birmingham with 1,002,603 people and Llanwrytd Wells Urban District in Brecknockshire, with 742. We have computed 1931 unemployment rates for the 376 modern local authorities via geography conversion tables created using our boundary mapping and 1931 parish population statistics (Simpson, 2002), and the results are shown in figure 1.

Figure 1: Census Unemployment in 1931 redistricted to 2001 Local Authorities



(darkest shade = unemployment over 15%)

Where the whole of a 1931 local authority fell in just one modern district, we can simply aggregate. However, many historical districts overlap with more than one modern district, so we have to allocate the 1931 data between them. Knowing the parish populations and boundaries makes this allocation more accurate, although we are of course assuming that the unemployed had the same geographical distribution within the 1931 district as the population as a whole, and that population was equally distributed within each parish. The assumption that the unemployed had the same distribution as the population can be avoided through modelling approaches which embody alternative distributions but, even if we could deduce the likely distribution of unemployment within districts, these techniques are hard to explain to a non-technical audience. Obviously, any attempt to create redistricted data for output units which were simply subdivisions of large reporting units such as Birmingham CB would be quite spurious. This is the main reason our results are presented for modern local authorities,

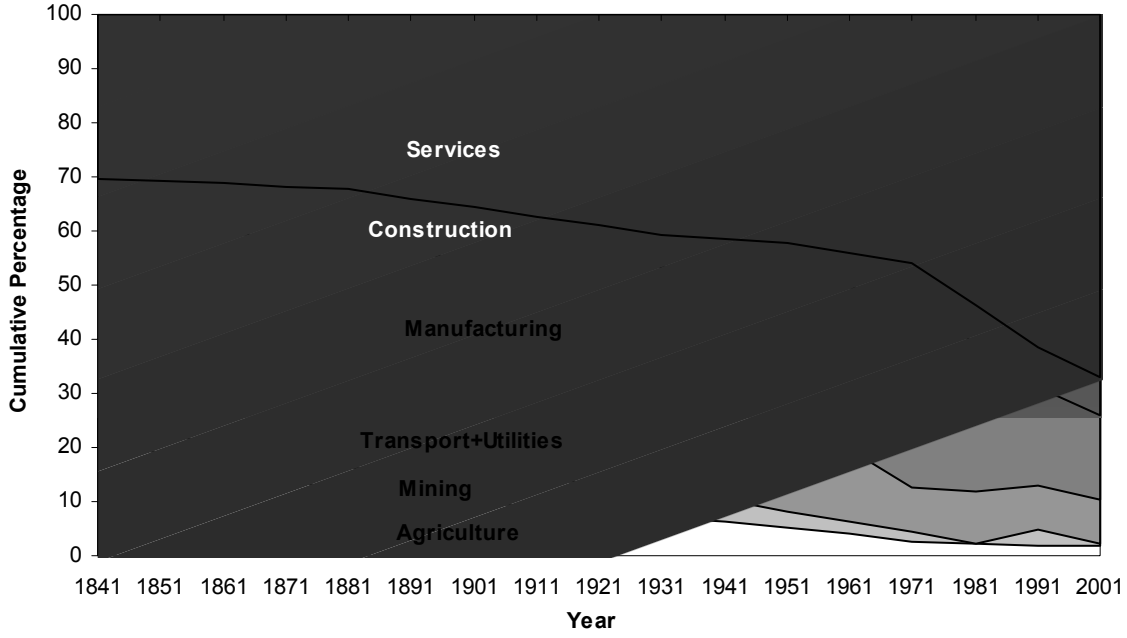
which generally cover even larger areas, rather rather than for Parliamentary Constituencies as advocated by Danny Dorling (1998).

Meanwhile, our statistical focus had broadened from unemployment and other measures of economic distress to include the census and mortality statistics. This resulted partly from collaborations, as significant funding had come through other projects passing on parts of their grants, but is was also reflected continuing research into the north-south divide. One strand concerned labour migration, the lack of which is used by standard explanations to explain the persistence of regional inequalities. Earlier work on trade union membership records had shown that nineteenth century skilled engineers were remarkably mobile. However, studying long-run migration trends in the wider population was difficult because the obvious census data was on place of birth versus place of current residence, so it was hard to say even in which decade a move occurred, and because flows could rarely be disaggregated below the county level. We developed an alternative method, ignoring the birthplace tables but combining age- and gender-structure data from consecutive censuses with similarly age- and gender-specific mortality data to estimate *net* migration flows via a cohort survival method.

The other strand needed detailed occupational data for localities to explore the ideas of John Langton and Doreen Massey on spatial divisions of labour. A quick summary, which both of them would probably disown, is that, contrary to conventional wisdom, the Industrial Revolution re-emphasised differences between areas precisely because improved transport technologies created national markets for goods. In pre-industrial times, each district had to be largely self-sufficient in food, in building materials, in clothing and so on, so every district had to contain some producers of each type of good. With better transport, localities could specialise. Some of these specialisations are obvious, such as mining towns and mill villages: in 1841, 38% of the occupied population of the modern district of Blaenau Gwent were miners, rising to 55% in 1931; 75% of workers in modern Blackburn with Darwen were in manufacturing, reflecting the fact that 63% of the old Borough of Blackburn worked in textiles. Although many will mourn the decline of mining and manufacturing, these figures clearly indicate an extreme shortage of all kinds of services for the local population. However, other communities were already highly focused on services: 51% of the workforce of Leamington Spa in 1841 were domestic servants. Services were already a large part of the economy of England and Wales in 1841, with 30% of all workers, and as they have expanded and as distinctive heavy industries have declined, the division of labour between places has come to be defined not through what people make, but through their differing roles in a single

production process that involves several locations: the senior managers are in London, the R&D staff in the rural south-east, the warehouse in the East Midlands and, if it is in Britain at all, the factory somewhere with cheap labour and subsidies for capital investment. Such patterns have obvious consequences for income levels, class structure and vulnerability to recession (Gilbert and Southall, 2000).

Figure 2: Changing Occupational Structure 1841-2001: England and Wales



Winning the lottery

Patterns like these are not hard to grasp if you can *see* them, especially if they are presented as a comparison between your own home area and the national average. However, as discussed in the introduction, the data have tended to be very hard to access. The project's earlier work had solved some of these problems: the data were computerised rather than parts of fading and disintegrating reports, and our boundary mapping and methodological work has eventually enabled us to provide clear comparative time series for consistent geographical areas. However, our work was available only via data archives which served only the higher education community, and in articles which tended to reach only a similar audience. In reaching a wider audience, part of our appeal was that we could provide information on everybody's own home area, but this would be hard to achieve in a traditional paper publication: including graphs for each of 408 local authorities, for example, would be both expensive and pretty boring. Similarly, most

modern thematic atlases are published as small format books with few colours; identifying particular districts would be almost impossible.

The obvious solution was electronic publication. The original 1989-91 database project had done some experiments with a simple system – to the user – that created interactive maps on a desktop computer by linking to a database over a network. We were a bit ahead of our time, as we were doing this before the web, and before British universities linked to the internet; our system used Apple Macintoshes running Hypercard software, and some specialised software to connect to our database that we persuaded Apple Computers to let us play with (Gilbert and Southall, 1991). In 1996, we got funding from the Joint Information Systems Committee (JISC) to do more experiments with on-line historical mapping, although as a higher education sector body they would never have funded an actual web-based system for the general public (Southall and White, 1997). Most research funding bodies, and certainly the Economic and Social Research Council, have clear rules excluding funding for publishing projects.

Who would fund something like this? At times, commercial publishers have shown a little interest, but few have made money out of on-line educational publishing, and the most successful have often been those who charged the most. Unlike the HE sector, where the JISC provides a communal infrastructure for all universities, schools and public libraries are generally funded through local government with no real central body. Some kind of 'lottery grant' was the obvious solution, and as already mentioned I applied to the Millennium Commission in 1996. They received a good many such IT-based applications then, but they preferred to concentrate on large and *very very useful buildings*, and only one IT project was funded, the Scottish Cultural Resource Access Network (SCRAN). However, following the 1997 General Election, the Heritage Lottery Fund's rules were changed to encourage interpretative projects – although we have yet to try out the idea of 'statistical heritage' on them – and a new lottery funding body was created, the New Opportunities Fund (NOF), with a brief to fund projects in health, education and the environment, but entirely through specific programmes each requiring government approval. In the autumn of 1999, NOF announced their 'Digitisation of Learning Materials' programme (NOF-Digi).

The programme as announced sought projects in three areas: 'cultural enrichment'; 'citizenship in a modern state', which included 'rights and responsibilities'; and 'reskilling the nation' which included numeracy and visual literacy (NOF, 1999, pp. 13-14). In principle, any

not-for-profit organisation, other than central government departments, could apply for funding under any of these heads. However, much of the detailed direction of the programme was the responsibility of Resource, like NOF a quango under the umbrella of the Department of Culture, Media and Sport. Resource's basic job is to advise and represent the interests of museums, galleries, libraries and archives – the heritage sector – and unsurprisingly publicity for the programme and many of its detailed procedures were targeted on that sector; the SCRAN project was presented as the main example of good practice at programme briefings. One result was that the vast majority of the final grants, which totalled £50m., were clearly targeted at 'cultural enrichment'. Despite the obvious potential links with the new 'citizenship education' strand of the National Curriculum promoted by the Department for Education and Skills under David Blunkett, under £1.5m. went to projects with any connection with citizenship, and most of this went to a 'Citizenship Past' consortium whose main projects were archive/library based. Literacy and numeracy were even less visible. Given the range of organisations that could have applied for grants, there were many wasted opportunities.

I wrote an outline application to the programme in January 2000, and the initial emphasis was firmly on presenting historical census information. We claimed, I think justifiably, to be addressing all three programme areas: the numeracy aspect was obvious, and understanding how society is changing is essential to an active citizenship. The application was originally made by the project, which we admitted had no legal existence but did have a management committee, a fair sized staff at two sites, and a substantial existing income from grants. Once created, our web site would be kept funded by subscriptions from schools and libraries, which was maybe unlikely but was the model followed by SCRAN. We had hoped that a full bid would be supported by a learned society with an interest in statistics, but partly because a key letter got lost at their offices we were unable to pursue this and the application had to become one from the University of Portsmouth, to which the project had moved at the start of 2000.

In April 2000, NOF told us they had decided to adopt a 'consortium approach', based on their assembling the consortia. In June 2000, we were told we had been short-listed, and soon after discovered that we were in a nine-partner consortium called 'Sense of Place (National)', led by the British Library, and also including English Heritage, the National Trust and, bizarrely, the National Horse-Racing Museum. We never did discover their role as they withdrew from the consortium before its first meeting, and by the submission of the full bids in January 2001 only three partners were left: the University of Strathclyde, the British Library and ourselves.

However, we were clearly now part of the heritage sector, and we were increasingly working with partners for whom the main goal of our project was *not* improving access to statistics. The British Library's own plans focused on parts of their collection which were generally place-specific – historic maps, prints and drawings of landscapes, even recordings of particular local dialects – but needed some overall spatial framework to bring them together. They saw us as providing this, and offered substantial assistance in return. While the relationship with the BL was essentially something imposed on us by the funding body, we also had a longer-term relationship with the archives sector, who needed a definitive 'authority list' of historical administrative units, partly to assist in cataloguing. Given that our historical statistics almost all relate to legally-defined units, and we had problems ourselves deciding whether slightly different place-names in statistical tables meant different areas, the archivists' goals were broadly compatible with ours, but this still gave the project a rather different image.

The whole of the second half of 2000 was spent working closely with British Library staff on our applications. The final GBH bid consisted of 44 pages of 'business plan', a 50 page application form (mostly the same information but in a different order), and 24 pages of job descriptions and CVs for existing staff. I was required to supply both a 'mission statement' and a 'values statement'. The latter read, in part, as follows; the italicised bits pick up on goals of the programme, but members of the Radstats mailing list may be able to spot some wording 'borrowed' from Ludi Simpson's suggested 'defining statement' for Radical Statistics:

> This project certainly involves *cultural enrichment* … but its own core is … concerned with *reskilling the nation by developing existing and new forms of literacy*: numeracy, graphicacy and map reading. Further, our aim is to make information about society and individual communities accessible to all: we believe that meaningful statistics can provide an insight into social change which can help the least powerful groups in society realise their full potential. We aim to create *informed citizens in a modern state*.

This preparation work was of course un-funded, and I was able to do it only because I was at that time on an ESRC fellowship whose goals were (hopefully) sufficiently closely related to justify this. I was reasonably confident we would get some funding, as the New Opportunities Fund had said they hoped to fund all short-listed projects, but we knew that the larger projects were likely to be cut substantially. I finally heard we had been funded in mid June 2001, almost 18 months after the original application, and our grant was £620,000, a cut of only £78,000 from

what we had asked for – but still meaning the budget and business plan of course had to be revised.

Serving the people

The new funding did cover some new computerisation of statistics, mainly on mortality, education, language and parish-level population counts, and work on mapping Scottish boundaries. However, the new digitisation work was primarily about contextualising what we already had: the texts of the *General Reports* of every census up to 1961 plus selected other reports; the texts of 19th century descriptive gazetteers; scans of historic Ordnance Survey one-inch maps, to help users place our boundary mapping in context. The real challenge, though, was how to design a database providing a single coherent framework for all this digital *stuff*, and a web site that both made sense to the general public and could handle really large numbers of users.

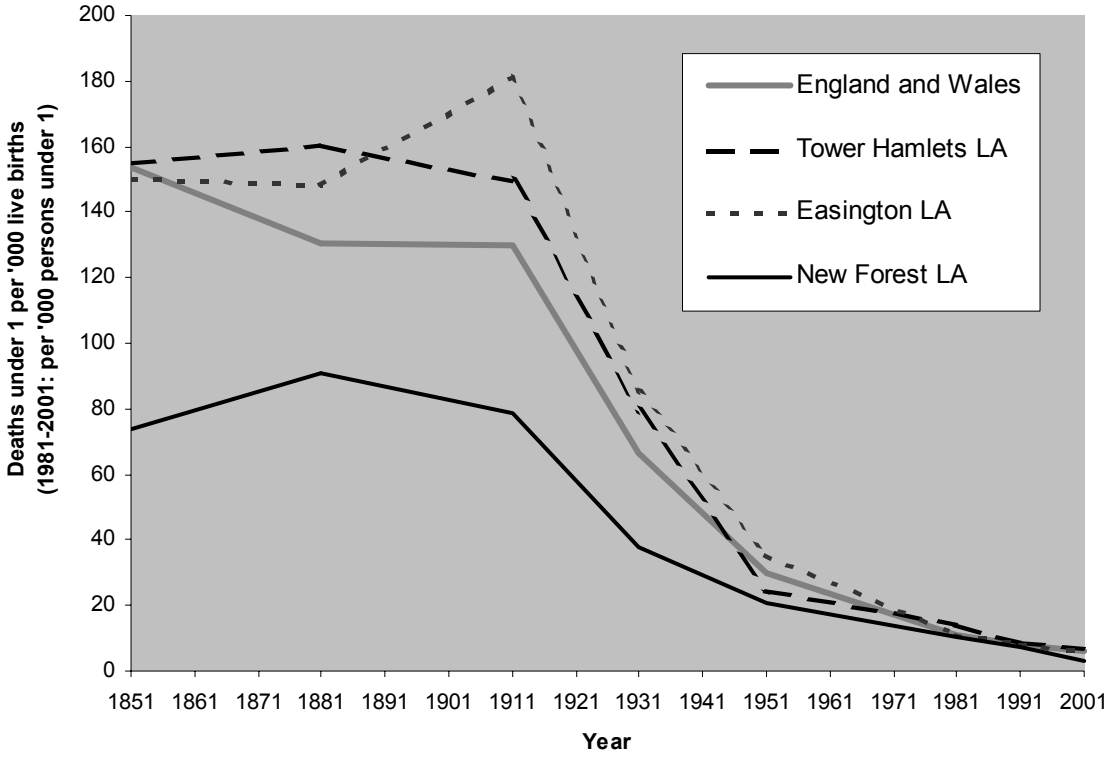
Most big on-line statistical systems are based on databases of tables rather than databases of individual statistics: they help users locate which table they should download, or maybe map. Our system is different: our users will be looking for data mainly on their home area, so they need to be able to select out just one or two relevant rows from each of the hundreds of published tables we have computerised. The way we do this is to store **all** our statistical data in just one database table, and in fact in just one column of that table. The size of this table is not itself a problem, as modern database management systems can happily handle tables with billions of rows, never mind the 30 million or so data values our final system will have. However, this approach means we are completely dependent on our metadata to locate data values of interest as, unlike text, images or even whole statistical tables, our individual data values are, obviously, almost indistinguishable. The current prototype holds 1,073,950 data values in the main data table of which, for example, 12,479 are the number '2' and 9 are '9,736'. Exciting!

Detailed technical presentations of the system will be appearing elsewhere. However, my broad approach is to record four different kinds of information about each number:

- **Source:** we have created a computerised inventory of all the British censuses up to 1961, all the reports published from each and, crucially, all the tables appearing in each of the reports. We can therefore hold source table IDs for each data value. These could be used in reassembling the original published tables, but we plan to also store our original table transcriptions as spreadsheets, which can far more faithfully reproduce things like column headings than any database system.

- **When:** the data table itself holds the date or period that each value covers.
- **Where:** in the original reports, this is recorded by the name of some county, district or parish, but these can be ambiguous – the same parish name may appear in several counties, while there are endless slightly different ways to type "St. Helen's" or "Newcastle-upon-Tyne". In a conventional Geographical Information System each number would be associated with a polygon, but we may never have digital boundaries for *every* unit ever listed in a British census report; and, as boundaries often changed, direct linkage to polygons means losing continuity. We therefore link each value to a unit ID value, based on our administrative units gazetteer which also provides the archivists with their authority list. That gazetteer has quite a complex structure, as each unit may change its name, its location, its status ('Urban District', 'Municipal Borough', etc) and its hierarchical relationships. Each of these is therefore held in a separate table, so units can be identified by any number of different names, including ONS codes. Locations are recorded by storing polygons within the database; we are using Oracle with its spatial extension.
- **What:** This is the hardest part, and our approach is based around a fairly new 'standard,' the Aggregate/Tabular Data Extension to the Data Documentation Initiative standard, which is available from http://www.icpsr.umich.edu/DDI. Very briefly, the DDI approach is to say that any published table is based on one or more *variables* in the original micro-data, i.e. a census schedule or death certificate; and each variable is divided up into a set of *categories*. A particular published table can therefore be described as a data matrix, or *nCube*, each of whose sides is one of the underlying variables. In our implementation, each cell in an nCube is identified by a *cellRef* ID code, and it is these identifiers which appear in the main data table. Some of these identifiers are self-explanatory, such as 'Tot_Pop', but this system enables to hold not just two dimensional data in a single column of a single table, but any number of dimensions. Three dimensional nCubes, such as age against sex against occupation or cause of death are quite common. In addition to the variables, categories and nCubes defined by the DDI, we have added the notion of a *rate* so the system knows, for example, that infant mortality is normally measured by the number of babies dying, not the number surviving, and it is normally expressed as a rate per thousand births, not as a percentage. (See figure 3.)

Figure 3: **Infant Mortality Rates 1851 to 2001: National totals and three selected 2001 Local Authorities**



To illustrate how this works in practice, consider the nine instances of '9,736' mentioned above. The first has the source ID 'EW1911AGE_M11', the date '1911', the g_unit value 11000241 and the cellRef 'AGE_100up_f_15_19': it comes from table 11 in the report on *Ages and Condition as to Marriage* of the 1911 Census of England and Wales; it covers the Poor Law Union of Gateshead, Durham in 1911; and it records the number of women aged 15 to 19. The second is the total population of Ilkley Urban District, Derbyshire, in 1931. The third is the overall total number of workers in the 1841 industrial data for the modern district of South Derbyshire, i.e. it is one of the results of our redistricting computations.

How does this structure help people find out about their local area? We assume that most people start with a postcode – exhaustive research within the project team has shown that everyone knows two such codes, their own and their mum's. Using standard Ordnance Survey CodePoint data, we convert this into a grid reference and then use our database's spatial abilities to do a *point-in-polygon* search of the boundaries in the system. That search will identify all the historical units containing the user's home, but we think most people will be confused by all the terminology – Rural Districts, Poor Law Unions,

Wappentakes and so on – so we take them first to the modern District or Unitary Authority they live in, with the option of selecting one of the historical units instead.

Our data documentation system assigns all variables, nCubes and rates to one of, currently, eight themes, such as 'Population' and 'Life and Death', each with an associated icon. Once the user has selected a unit, they can choose between the available themes and are then taken to a page which graphs each of the rates defined for that theme. We have used our re-districting techniques to convert historical data from every census, plus vital registration data from as far back as 1851, to the modern districts, so for these units we can present time series graphs covering 150-200 years. As well as data for their own area, we always present the national rate for comparison; we hope to add an option to change the comparison unit so that two districts can be directly compared. These graphs are accompanied by a commentary on each rate, although one obvious limitation is that we cannot write text specific to each geographical area, apart from a basic summary of statistical trends 'written' by a computer program.

From these pages covering all the rates within a given theme for their chosen unit, users can go to a national map for a selected rate and date; for example, male unemployment in 1931. These pages can also be reached direct from the main home page by clicking on a thematic icon rather than typing in a postcode. Users can zoom in and out to see additional detail in each map, or they can click on a particular unit to move to the relevant thematic page for that area. Another option is to change the geographical units used in the map. For example, if the current map shows the proportion of the whole population aged over 65 in 1911 for modern local authorities, the user can request alternative maps of the same rate for either Registration Districts or Local Government Districts, as the 1911 census reported age structure for both. Another option from the pages for themes for a given unit is to go deeper into the statistics, exploring the available nCubes.

One major problem for the project would be if no-one wanted to use our web site. Another possible problem, however, is how to cope with all the amateur local historians in Britain wanting to use our site, especially in the weeks when all the nine year olds are trying to prepare the local history projects required by the English National Curriculum. The problems the Public Record Office had with their 1901 Census web site served as an awful warning for us, but it is clearly impossible to design a web site, especially a highly graphical and interactive one, so it can handle any conceivable level of use. What we have done is make sure that the site could be fairly easily scaled up to serve a larger audience by simply – if expensively – buying additional server computers. At present,

the whole system runs on just one server, but it consists of several separate programs which could each run on a different computer.

Almost all the web pages in our system are created by software in response to requests from users. The basic pages are defined by programs written as Java Server Pages (JSPs), running within an Apache web server using the Tomcat servlet container. However, all the graphics within the pages are generated by specialised servers: the graphs by Cewolf, the thematic maps by GeoTools and the scanned maps, which we use as an underlay beneath our boundary lines, by MapServer. All of this software is 'open source', meaning it is non-commercial and 'free', but all of our data apart from the scanned maps are stored in an Oracle database, which is commercial and decidedly expensive. The basic web server could run on one computer, the underlying database on another, and each of the graphics servers have its own computer.

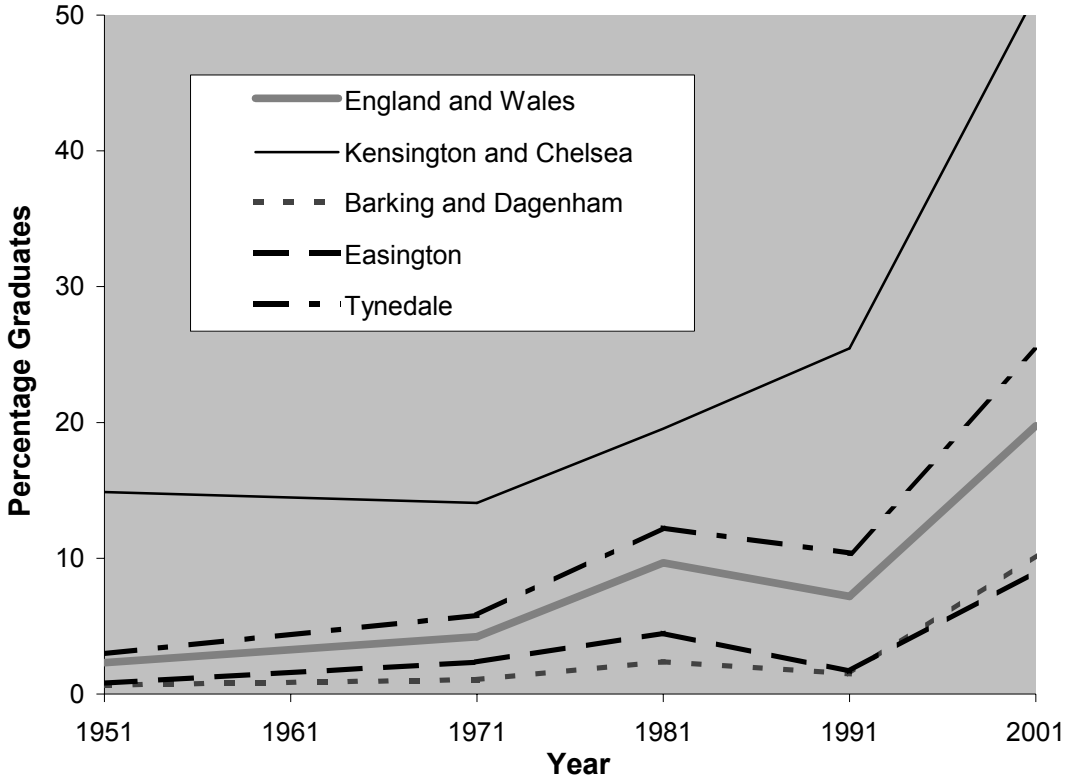
And it all ended ...

Does 'statistical heritage' matter? And is the world ready for it?

Almost all statistical work uses statistics that were not gathered the same day, so all distinctions between 'modern' and 'historical' data are matters of degree – and, in practice, census data back to 1971 seems often to be viewed as 'modern' simply because it is available in computerised form. Further, trends in many aspects of society can only be understood if we examine data for many decades. This is very obvious indeed for demographic statistics but is also true of many measures of economic and social structure, which change fairly slowly. My earliest work was about gathering unemployment data from the pre-1914 period, and showing that the resulting geography was not that different from 1931 – or 2001.

Figure 4 presents some of the education data we have assembled: despite a vast expansion in university education over the last 50 years, districts have changed their rankings by proportion of graduates relatively little.

Figure 4: University Graduates 1951 to 2001: National totals and four selected 2001 Local Authorities



A more specific demonstration of how historical data can be given a modern application is an analysis which uses the historical GIS to examine how today's elderly were affected by their childhood in the inter-war Depression. The ONS Longitudinal Study includes information on whether its members – about half a million people – died between 1981 and 1991, and whether the survivors were experiencing 'limiting long-term illness' in 1991. It also identifies its members by National Health Service number. If you were alive on September 30th 1939, and did not subsequently serve in the armed forces, your NHS number is actually the National Registration number issued that day, and it identifies the local authority you were living in. We were therefore able to link modern data about individuals from the Longitudinal Study with 1931 data on local authorities, such as unemployment rates and class structure. The results show that, even allowing for many aspects of their more recent experiences, and regardless of where they have since moved, people brought up in areas of high unemployment in 1931 have significantly worse health today (Curtis *et al*, 2003).

It is less clear that the world sees this stuff as important:

- The National Curriculum does require every schoolchild in England to prepare 'A study investigating how an aspect in the local area has changed over a long period of time, or how the locality was affected by a significant national or local event or development or by the work of

a significant individual', using information on 'education; population movement; houses and housing; religious practices; treatment of the poor and care of the sick …' – mostly statistical themes. The National Curriculum web site (http://www.nc.uk.net) provides links from each requirement to relevant web resources, but at present the links from the 'local study' requirement are all to methodological guides or information about particular localities, not sources of local information with national coverage – and they include examples of completed census forms, for example, but no statistics derived from them.

- The Office of National Statistics have become increasingly interested in historical data. However, although they have given us strong support, they are not in the business of long-term preservation. The archiving of official statistics, like all other government records – as distinct from publications – is the responsibility of the National Archives (until recently the Public Record Office). A meeting on "Archiving Statistics: Challenges and Prospects" at the Royal Statistical Society on 1998 established their interest but practical results are unclear. One particular frustration is access to the 1961 census microdata: magnetic tapes containing these data were passed to the Archive more than three years ago, and could be used to create new small areas statistics or anonymised records, but remain unused.
- Although our original application for lottery funding was very much about improving access to historical statistics, we got funded because our work on place-names and boundary mapping was seen as relevant to more conventional heritage sector content. As the British Library put it in 2000, 'we see [the GBHGIS] approach as being a facilitator for access to our own material'. Unfortunately, the web site they have eventually developed is mainly organised around the different collections within the library rather than by geography, and it is now unclear who will meet the long-term running costs of our site –a familiar tale with lottery-funded projects, but obviously very frustrating for us.

Lastly, how do you get to see all this stuff? After many delays, the first version of the site 'went live' in September 2003, with more content being added in stages over the next twelve months (for example, the initial system did not contain any mortality data but the next release does). The address is:

www.VisionOfBritain.org.uk

References

Curtis, S.E., Southall, H.R., Congdon, P. and Dodgeon, B. (2003), 'Area effects on health variation over the life-course: analysis of the longitudinal study sample in England using new data on area of residence in childhood', *Social Science and Medicine*, in press.

Dorling, D. (1998) 'Is it worth reducing inequalities in health?', *Radical Statistics*, 68.

Gilbert, D.M., and Southall, H.R. (1991) 'Data Glasnost: a user-friendly system for access to research databases across wide-area networks', *History and Computing*, 3: 119-28

Gilbert, D.M., and Southall, H.R. (2000) 'The Urban Labour Market', pp. 593-628 in Daunton, M. (ed.) *Cambridge Urban History of Britain, vol.III: c.1840-1950* (Cambridge: Cambridge University Press)

Johnson, P. (1993) 'Small debts and economic distress in England and Wales, 1857-1913', *Economic History Review*, 2nd. series, 46: 65-87

New Opportunities Fund (1999) www.nof-digitise.org: *Information for Applicants* (London, NOF)

Simpson, L. (2002) 'Geography conversion tables: a framework for conversion of data between geographical units, *Int. J. Population Geography*, 8: 69-82

Southall, H.R. (1988) 'The Origins of the Depressed Areas: Unemployment, Growth, and Regional Economic Structure in Britain before 1914', *Economic History Review*, 2nd series, 41: 236-258

Southall, H.R. (1998) 'Working with historical statistics on poverty and economic distress', in D.Dorling and S.Simpson (eds.), *Statistics in Society: The arithmetic of politics* (Arnold), pp.350-8.

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*, 2nd series, 49: 35-57.

Southall, H.R., and White, B.M. (1997) 'Visualising Past Geographies: the use of animated cartograms to represent long-run demographic change in Britain', pp. 95-102 in Mumford, A (ed.) *Graphics, Visualisation and the Social Sciences* (Loughborough: Advisory Group on Computer Graphics, Technical Report Series no.33).

Humphrey Southall
Great Britain Historical GIS Project
University of Portsmouth, Department of Geography
Buckingham Building
Lion Terrace, Portsmouth PO1 3HE
Telephone: 023 9284 2500
Humphrey.Southall@port.ac.uk