

Applying historical GIS beyond the academy:

Four use cases for the Great Britain HGIS

Humphrey Southall (University of Portsmouth)

Introduction

Many historical GIS projects are the work of individual scholars, carried out in their own research time without external funding. Most of the projects which do receive external funding are relatively small scale, employing a single research assistant to work alongside the principal investigator. However, a small minority of historical GIS research projects are among the most expensive projects of any kind in the arts and humanities. They are also more expensive than most non-historical academic projects using GIS technology. This is because the latter can use the vast bodies of geo-referenced data describing the modern world that are available from national mapping agencies and through remote sensing. Conversely, even where historical maps are available the historical researcher needs to scan, geo-reference and probably vectorize them; and often spatial data need to be constructed from textual information containing geographical names, not coordinates.

As a result, it is hard if not impossible to fund a major national historical GIS, one that identifies every town and village, and covers a century or more of change, with the kind of funding available for academic historical research. Total funding for the Great Britain Historical GIS now totals over \$US 3.5m., but obtaining this funding required us to demonstrate that the results would benefit an audience beyond academic history. This chapter describes how we have worked with four other groups to meet their needs: health researchers, archivists, government environmental agencies and companies selling advice to the property sector.

Four caveats are needed. Firstly, the main focus is on the reasons for the collaboration, with details of the actual research provided mainly via references. Secondly, be very clear that our main source of funding remains grants, not commercial contracts, but the kind of evidence presented here of wider economic and social benefits has been crucial to obtaining grants. The other benefits of this kind of activity are further discussed in the conclusion. Thirdly, the case studies presented here are of course all drawn from the experience of one project, but that experience is quite diverse, and each case study ends with some more general lessons.

Possibly the largest caveat is that the case studies here do not include much the largest audience for our work, the users of our web site *A Vision of Britain through Time*, as originally funded by the UK National Lottery. Courtesy of Google Analytics, we can supply impressive statistics of raw volume, such as that the site was used by 1,405,398 different people (“unique users”) in the year from July 2011 to June 2012. However, because the site is completely open access we know relatively little about who these people are, and can only infer their motives. The available usage data will be further analyzed in the third of three papers on our rebuilding of the Great Britain Historical GIS,¹ but the focus here is on professional audiences we have worked closely with.

Demography and Health

Most national historical GIS projects at least started with the goal of providing a framework for the analysis of historical census and vital registration data. This is certainly true of both the Great Britain Historical GIS and the US National Historical GIS. Both projects not only created computerized boundaries for the main demographic reporting units, they also assembled large bodies of historical statistics. However, my aim here is not to review the very clear contribution of historical GIS to historical demography, but to explore what more needs to be done to make it useful to non-historical demographers, and especially

to contemporary medical researchers; and, more importantly, to explain *why* historical GIS can make a large contribution to modern medical research.

Twenty years or so ago, the growing number of retired people, especially those aged over eighty, produced near panic among policy makers in advanced societies, as they were seen as an inevitable large burden on health systems as well as pension funds. It is now recognized that many people have a healthy old age, requiring constant care only in the final months of life, but this has led to a new emphasis on research into the factors deciding who will experience healthy old age, and how the proportion of such people can be increased. This has become linked to medical research which has demonstrated a strong relationship between what happens to people before they are born and in infancy, and their health much later in life. This research began with a classic study by Barker into the relationship between a baby's nourishment, as recorded by weight at birth and subsequently, and the risk of death from coronary heart disease.² More recent research into the "Barker hypothesis" has explored the wider impact of deprivation in infancy.³ Such research forms the background to major government programs such as Head Start in the US and Australia, and Sure Start in the UK.⁴

Given that such research is an urgent public priority, starting gathering data on babies now and reporting the results in sixty to eighty years time will not do: it has to be based on recent data on health in old age for individuals whose experience as babies has already been recorded: it has to be based on historical data. Three different approaches have been taken in UK research.

Firstly, Davey Smith and his collaborators located data on 4,999 babies studied by Boyd-Orr in 1937-9, which included, as well as measurements of height and weight, names and addresses. The same individuals were then in 1988 located for them in the modern National Health Service Central Register and asked if they were willing to take part in a follow-up study.⁵ Note that all the research described in this section is substantially limited

by rules governing access to confidential data on living people and their health, although space does not allow a detailed account.

Secondly, Strachan used data from the Office of National Statistics Longitudinal Study (LS) which links together data for one per cent of the population of England and Wales from the 1971, 1981, 1991 and 2001 censuses, plus information on events such as births, deaths and cancer registrations; it covers just over half a million people at each census.⁶ That clearly tells us nothing about conditions in childhood, but Strachan discovered that the members of the LS were identified by their National Health Service identity numbers and that, if they were alive on the 30th September 1939 and did not subsequently join the armed forces, their NHS number was the National Registration number they were issued on that date, which identified the local authority area they were then living in. Strachan showed that LS members who were in the north of England in 1939 had worse health today, even if they were now living in a healthier region.⁷

This is where the GB Historical GIS first became involved, as Strachan's work was based on some broad assumptions about certain regions being deprived, not on historical data about actual localities. We assembled a range of data on unemployment, social class and overcrowded housing from the 1931 census, and contemporary infant mortality data, from the 1931 census and supplied it to the LS support team, who made our data a permanent part of the LS. The LS already contained two variables measuring individual health: whether or not the individual died between 1981 and 1991, and whether or not they reported in 1991 that their lives were limited by long-term illness. We then ran new analyses which related these two outcome variables to both individual-level data on LS members' recent experience and the locality-level data from the 1930s we had contributed.

Two factors greatly complicated this work. Firstly, we had no direct access to the LS, instead supplying data to the LS support unit and specifying analyses for them to run.

Secondly, although the basic architecture of local government remained constant between 1931 and 1939, consisting of County Boroughs, Municipal Boroughs (all urban units), Rural Districts and London Boroughs, the detailed geography of local government was greatly changed through a rolling program of county reviews: the 1931 census that provided most of our explanatory variables reported on 1,800 Local Government Districts, but this had been reduced to 1,472 by 1939. Further, many of the districts which were not abolished were altered through boundary changes: 289 (19.6 percent) of the 1939 districts were new creations or had been affected by boundary changes. To solve this problem, we constructed a geography conversion table⁸ from the 1931 and 1939 reports plus the 1,805 boundary changes listed for the intervening period, using 1931 populations rather than geographical areas. By linking this table to 1931 census data, we cut the data for the 1,800 1931 districts up into 2,916 fragments, and then reassembled them into the 1,472 1939 units. The table was very carefully cross-checked by comparing the 1931 populations of 1939 units computed by applying the boundary change information to the 1931 census figures, with the 1931 populations listed in the 1939 report. The results showed that people brought up in Britain's depressed areas in the 1930s have significantly worse health today, even if they later moved to other localities.⁹

The third approach works with samples of people, inevitably small, who have been surveyed repeatedly starting at birth. The first such British project, the National Survey of Health and Development, covers 5,362 people born in March 1946.¹⁰ Similar "birth cohorts" were started in 1958, 1970 and 2000, but the NSHD is the most relevant to aging research as its members are now in their mid sixties. Our work on the NSHD has two distinct elements. One task was assisting in converting postal addresses for cohort members into geographical coordinates. The 3,354 addresses from 1999 all include postal codes and locations were easily found for all members. All but fifty of the 3,519 addresses from the 1972 survey were

automatically matched against a modern address list. However, of the 4,856 addresses from 1950, 2,658 (54.7 percent) could not be matched automatically. Although the number of residential addresses today is vastly greater than in 1950, the majority of houses and flats that existed in 1950 still exist, so many of the non-matches were due to detailed variations in spelling and punctuation. However, our particular focus was the relatively small fraction of 1950 addresses which simply no longer exist, because these were concentrated in areas in which the street pattern has been completely replaced by “comprehensive redevelopment”. These are inner city areas with very poor housing conditions, whose consequences the project is especially interested in. Locating these addresses involved working with geo-referenced scans of mid-twentieth century mapping at 1:10,560 scale.

The other focus of our work is including additional area-based data in the study. There are two distinct reasons for doing this. One is that when the NSHD was started it was more narrowly medical than today, so while we know a lot about the babies’ physiology we know little, for example, about the houses they lived in. Aggregate data from the 1951 census about, for example, the percentage of homes lacking their own toilet can serve as a proxy for data on whether or not the cohort member’s home had a toilet. The other reason is that some factors, such as pollution, are inherently area-based. For example, in an analysis of the determinants of individual-level infant mortality as recorded by the special fertility survey within the 1911 census, Garrett *et al* found that children in certain types of occupationally specialized community had higher mortality rates, even when fully controlling for parental occupations: the children of teachers in mining communities were more likely to die in infancy than the children of teachers elsewhere.¹¹ Our analysis of the NSHD has already shown that growing up in areas with a high percentage of the population in a low social class significantly impacts physical capacity in later life.¹²

Our work with the NSHD is continuing, and means we are working closely with geneticists and historians of nutrition. Directly parallel research may not be possible in all countries, as it depends on the existence of these major individual-level longitudinal studies. However, a general precondition for historical GIS to contribute to medical or demographic research, as distinct from the history of medicine or historical demography, is that it must construct a continuous narrative of geographical change coming up to the present. Most British historical demographers, including those using GIS-based techniques, stop in 1911 when demographic reporting stopped using the relatively simple system of Registration Districts.¹³ Conversely, much of our work has focused on the subsequent system of local government districts used up to 1974, and our data holdings come right up to the present. Our work on the Longitudinal Study was possible only because we had reconstructed the very complex history of changes to local government districts during the 1930s.

Archives and “name authorities”

It is unlikely that any academic seeking to develop lucrative lines of contract research would choose the archives sector, and few archivists would claim much knowledge of GIS. However, the largest single grant for the development of the GB Historical GIS, from the UK National Lottery, was possible only through support from the archives sector. Our first steps towards building a multi-national historical GIS were funded by the European Union’s Framework Program 6, but here again it was essential that our partners in the QVIZ project included the National Archives of Estonia and of Sweden.¹⁴

With the obvious exception of archives with map collections, archives have very limited interest in coordinate data, but they have a large interest in geographical names, and especially the names of administrative units. One consequence is that archivists are one of only two professions within the cultural heritage sector to systematically catalogue by geography. Archaeologists systematically catalogue finds by location, but are well-served by

mainstream GIS software and will not be further discussed here. In a systematic survey in 2009 of British archives by the Name Authorities and Indexing Working Group, part of the UK Archives Discovery Network (ADN), out of 84 archives replying, 77 percent said they maintained some form of place-name index, and most maintained one covering individual files or documents, not just broad collections.¹⁵

Unlike libraries, archives hold unpublished documents, generally created by organizations rather than identifiable individuals. The single most important catalogue information for an archival document is therefore the identity of the organization which created it. National archives are primarily concerned with the records of national governments, and their interactions with other nations. In the UK at least, the best organized *system* of archives is the network of local record offices. Local government bodies are required by law to maintain an archive containing, as a minimum, their own most important records, and one of the functions of the UK's National Archives is to inspect these local record offices, which generally speaking exist for counties and for the largest cities, other districts depositing their records with the relevant county office.¹⁶

In other words, archives catalogue their records by the organizations, or “corporate bodies”, which created them; but the most important corporate bodies they are concerned with are national and local governments which are defined by and usually named after the geographical areas they govern. Further, in most countries and certainly in Estonia and Sweden as well as the UK, much the largest group of archive users are not academic historians or government officials but ordinary people researching the history of their families. As well as being the most numerous group of users, family historians are economically important to archives because they are willing to pay significant sums for facsimiles of archival documents concerning their families, such as birth certificates and wills. These documents were generally created by historical administrative units such as

parishes and Registration Districts, and amateur researchers may have large problems identifying which administrative area contained their ancestor's place of residence; in other words, they need to identify the polygon containing a known point. One key recent development is that an increasing proportion of archival documents, especially those of interest to family historians, have been digitized and are available for purchase on-line, so researchers do not need to visit the places their ancestors lived in but they still need to find documents organized geographically.

Traditional archive catalogues, like traditional library catalogues, are held on filing cards arranged alphabetically. Computerizing these catalogues has usually meant transferring the text from each card to a database record. In this context, creating a "geographical catalogue" means associating *one* geographical name with each record. One problem is deciding which name to use; for example, the name of a house, the name of the street it was in or the name of the settlement. English local record offices have tried to standardize on using the name of the parish. The second problem is deciding the exact form of the name, which may have large consequences for arranging the cards in alphabetical order. For example, "Great Barford", "Barford, Great" or "Barford Magna"? Should the saint's dedication of the parish be included, and if so where: "Barford St. Margaret" or "St. Margaret Barford"? "Great Barford" is fairly clearly a different parish from "Little Barford", but is it the same as "Barford St. Margaret"?

"Name authorities" are lists that define which form of a name should appear in a catalogue. Where the catalogue is computerized, name authorities define a "controlled vocabulary", meaning the set of character strings which can appear in a given database column. Although a name authority can be a simple list, a better name authority will include "alternate names" and identify for each which "preferred name" should be used instead. In Wales or Estonia we also need to identify the languages names are in. A name authority of

the parishes of England will also need to deal with the many parish names which correctly refer to more than one parish, names such as “Newton” and “Aston”. The usual way to *disambiguate* such cases is to identify higher level units, most obviously counties, that contain the parish. A reference work that includes preferred and alternate versions of terms, and places them in a hierarchy, is a *thesaurus*.

By now, anyone with a GIS background will be wondering what the difference is between a name authority and a gazetteer, and anyway why not use coordinates rather than names, removing all problems of locational ambiguity; only three out of sixty-two archives in the ADN survey included a coordinate in their indexing of places. There are three problems with coordinates. Firstly, archival documents rarely contain them, so even if we are going to store coordinates in the catalogue we still need a reference resource that translates the names in the documents to coordinates. Secondly, no human being could search a card catalogue containing coordinates, and the off-the-shelf records management packages used by most archives, like Calm and Adlib, are modeled on card catalogues and completely lack spatial functionality; and no GIS software in existence provides the broader document management capabilities which archives require. Thirdly, while one aim clearly has to be to raise awareness of spatial functionality among archivists and their software suppliers, there are real performance problems with adding spatial functionality to heavily used on-line systems, while name-based searching is computationally very efficient. Conversely, it is impossible to build a digital name authority on top of packaged GIS software, as there is no mechanism for holding multiple names or labels for each entity. While card indexes may be going out of fashion, most people looking for place-specific information type place names into Google, and basing our web site *A Vision of Britain through Time* around an ontology-based architecture rather than a conventional GIS has consequently led to much higher usage.¹⁷

Obviously, a gazetteer can be a name authority, especially if it includes variant forms as cross-references to a main entry; several British local record offices use *Bartholomew's Gazetteer of the British Isles* as their main name authority for place names. However, when the UK National Council on Archives (NCA) published their *Rules for the Construction of Personal, Place and Corporate Names* in 1997, they prioritized a set of books listing administrative areas: F. Youngs' *Guide to the Local Administrative Units of England* (1979 and 1991), M. Richards' *Welsh Administrative and Territorial Units* (1969), and the General Register Office's *Index of Scottish Place Names from 1971 Census* (1975).¹⁸ One oddity revealed by the ADN survey is that of the 52 archives who responded to the relevant question, 32 said they were following the NCA Rules on place-name indexing rather than an in-house standard, but the two commonest authorities cited were Bartholomew's *Gazetteer of the British Isles* and the Getty Information Institute's *Thesaurus of Geographical Names*, both of which cover "places" rather than the administrative units emphasized by the NCA Rules.

Although our national lottery funding required us to create a public web site with wider popular appeal, a central deliverable was a computerized replacement for the authorities identified by the NCA. This clearly could not be achieved by modifying our existing ArcGIS system. Firstly ArcGIS could not hold a mass of variant names for each entity, plus additional attributes each name. Secondly and more fundamentally, much the largest of the authorities identified by the NCA was Youngs' *Local Administrative Units*, and this contains no maps or coordinates; and we clearly lacked the time to research boundaries for all the units Youngs lists. Conversely, Youngs provides a mass of textual information on relationships, and especially on hierarchical relationships: district A is in county B, and contains parishes C, D, etc. We had therefore to develop an architecture in which hierarchical

relationships were required but boundary polygons were a highly desirable extra; and in which entities could have any number of names.¹⁹

[Figure 1 near here]

Figure 1 uses the parish of Carisbrooke on the Isle of Wight to illustrate the resulting architecture, and in particular the organization of the system around hierarchical relationships. While this structure could be implemented in any relational database, we actually implemented it in an *object*-relational database, originally Oracle and now Postgres. We could then use the object extensions to hold the boundary polygons from the original system as unit attributes. Because the core of this system is a set of entities rather than a set of terms, and because it holds more than one kind of relationship, it is not just a thesaurus but an ontology. Our current production system defines 79,269 units, 129,697 names for those units, 251,261 relationships between these units. It includes 83,706 boundary polygons, but only 40,626 units have associated polygons: units have multiple polygons because of boundary changes.

[Figure 2 near here]

While figure 1 shows something very different from a traditional GIS, the system can do far more for units which have associated polygons. Figure 2 shows the interface developed to our data structure by the QVIZ project, which allows non-expert users to identify historical units of interest via a point-in-polygon query. The QVIZ system then provided access to archival documents from the selected unit.

What is relevant here is that this system should be seen as not simply a response to the specialized needs of archivists. It provides a better foundation for a popular on-line resource for local history than any mainstream GIS. Even more importantly, it arguably provides a better way forward than a conventional GIS in organizing large bodies of demographic data for historical research. Ultimately, archivists, family historians and historical demographers

share a strong interest in historical censuses and vital registration data, and the geographical units used for reporting them. In England that primarily means parishes and Registration Districts, and it is not a coincidence that the main units of interest to the National Council on Archives had already been mapped by the GB Historical GIS with funding from research bodies. Similarly, the QVIZ project was led by Sweden's Umeå University and drew on work by their Demographic Database.

However, working with archivists, or the heritage sector more generally, does mean basing your work around genuinely open standards for data and metadata. Digitization funding programs come with strict rules on standards compliance, and it is possible to avoid following widespread standards which provide inadequate support for geography, such as Dublin Core²⁰ and Encoded Archival Context,²¹ only by knowing about and implementing more appropriate alternatives, notably the work of the Open Geospatial Consortium.²² Claiming that Shape files are a “*de facto* industry standard” will not do.

Lastly, while this section has discussed working with archivists, with family historians sometimes the ultimate beneficiaries, the underlying aim of this work is to make the world's archives searchable by geography. This would be of immense benefit to academic historians and especially historical geographers.

Environmental Management

One regrettable oddity is that although many environmental historians now make significant use of GIS technology, this work has developed separately from “historical GIS”. One reason is that environmental historians have focused on raster data while historical GIS, so defined, has focused mainly on vector data, and on tools like ArcGIS and MapInfo. However, some of the largest technical challenges in historical GIS concern the adaptation of image processing technology developed for the analysis of aerial and satellite photographs to the automated analysis of historical maps.

We became involved in this area through work on historical land utilization survey maps, developed in collaboration with and partly funded by the UK's Department of the Environment, Farming and Rural Affairs (DEFRA), and two of its executive agencies, the Environment Agency and Natural England. One interesting aspect of this work is the differences in goals and perspectives between even these very closely related bodies.

Firstly, DEFRA's Agricultural Change and Environment Observatory was established to monitor, and where possible anticipate, changes in agriculture, particularly those arising from the 2003 Common Agricultural Policy reforms, and to assess the environmental implications of these. Their main focus is on strategic trends, rather than managing particular local environments. One issue is how changes in the structure of farm businesses affects farming practices, and so changes the landscape.

Secondly, the Environment Agency is concerned primarily with rivers, flooding and pollution, and is consequently focused on the management of river basins. It has primary responsibility for implementing the European Union's Water Quality Directive. One aspect of this is reducing the amounts of nitrates, "nitrate vulnerable zones" currently covering 20 percent of England. Given that reducing fertiliser use may make arable farming non-viable, identifying which areas and additional farms were traditionally pastoral and converted to arable during or after World War II is one way of targeting action.

Thirdly, Natural England exists "to ensure sustainable stewardship of the land and sea so that people and nature can thrive. It is our responsibility to see that England's rich natural environment can adapt and survive intact for future generations to enjoy."²³ One aspect of this is the Environmental Stewardship scheme, funded by the European Union as part of the Common Agricultural Policy, Pillar 2. UK funding under this for 2007-13 is £4.5bn. "Place based evidence of about long term change will greatly aid in our interpretation of the evidence arising from over landscape monitoring programmes such as Countryside Quality

Counts”.²⁴ One way this can happen is through better understanding of the landscapes associated with “low input farming”, which we are maybe moving back towards.

All of these agencies therefore have an interest in past land use, and especially agricultural land use. To some extent, any conventional topographic map provides information on the uses people make of the land they live on: they usually identify buildings and transport routes. However, large areas of topographic maps are simply empty, or identified only as fields. In recent years, we know considerably more from satellite imagery, especially “false color” images showing non-visible wavelengths. Some information for historical periods can be derived from sources such as tithe maps but these were not created for all areas and not all survive. For Britain, the first systematic national survey of land use was the Land Utilisation Survey of Great Britain, directed in the 1930s by Professor Sir L. Dudley Stamp of the London School of Economics (LSE), working in collaboration with the Geographical Association and county education committees; and through them with the country’s schools.

[Figure 3 near here]

Participating schools were supplied with instructions and with six inch to one mile maps of their locality. Pupils then went out into the countryside, supervised by their teachers, and recorded land use by marking each plot on the maps with a code letter, as shown in figure 3. These field survey maps were returned to Stamp’s team, and used to prepare one inch to one mile maps for publication. Stamp paid the Ordnance Survey for the use of the printing plates for the black layer for the then-current *Popular Edition* maps – effectively a GIS coverage – and then over-printed land use information as a series of different colors, including brown for arable, green for pasture, yellow for rough grazing and red for “agriculturally unproductive”, which covers both dense urban areas and industrial sites. Although the bulk of the survey was carried out quite quickly between 1931 and 1934,

funding publication of the maps proved much harder, even with government support once Stamp became Chief Adviser to the Ministry of Agriculture in 1942: a total of 167 maps were published between 1933 and 1949, using nine different printers with consequent variations in the inks used, and a further 56 maps showing upland Scotland were very carefully colored using water color paints and deposited with the Royal Geographical Society library in London.

Our first problem was establishing who owned copyright in the maps. The Ordnance Survey base maps had clearly been in Crown Copyright, but this lasts for only fifty years. An initial assumption was that the LSE owned some rights, but in 1936 they made Stamp sign an agreement taking full personal responsibility for the LUSGB. He was a very successful author of geography text books so his personal resources were significant, and he established and owned a company, Geographical Publications Ltd, which published the maps. It is absolutely clear that Stamp was the principal author of the maps, and copyright therefore runs for seventy years from his death in 1966. He left the company to a member of his team, with whom we have established contact. They have been enormously supportive of our work.

A second problem was assembling and scanning a complete set of good quality copies of the maps: our university had some, the Environment Agency bought a significant number from a dealer, and the LSE was able to sell us some mint-condition maps from unsold stock. However, the most important sources were other university map libraries, who freely lent us maps from their collections for scanning. Similarly, we were eventually able to include the unpublished maps of upland Scotland at minimal cost: a commercial re-publisher of historical maps was scanning a large number of other maps at the Royal Geographical Society and did the LUSGB maps for us, while the RGS also imposed no charge.

Once all the maps had been scanned by ourselves or our partners, geo-referencing and the construction of a seamless mosaic were time consuming but straightforward. The only

real complication was that the geo-referencing had to be done without using modern Ordnance Survey copyright data, as otherwise the OS would have been able to claim rights in the final product. However, our collection already included a digital version of the late 1940s *New Popular* edition of one-inch maps, which was the first edition to include the modern National Grid, making it trivially easy to geo-reference. We geo-referenced the LUSGB map scans by overlaying them on the relevant *New Popular* scans.

The large remaining problem is the systematic extraction of vector features from the geo-referenced images, and this work is continuing. One way of doing this would be to manually construct polygons by tracing around each color zone on each map, but this would be very time-consuming and expensive, costing upwards of £200,000. The alternative approach is to use image processing software, in our case Erdas Imagine, to automatically identify areas of similar colors. However, a number of problems with the maps complicate this: their complicated publishing history means that the original inks used to indicate a particular land use vary between map sheets, which means each sheet must be manually calibrated separately; the place-names and other information from the black base layer confuses the software; and a large problem especially in upland areas is that the reddy-brown used for contour lines is indistinguishable from the red used for “agriculturally unproductive”. Although solutions have been found for these problems, all require additional manual input and, to date, three separate technical studies have been funded by the Environment Agency to try to find the best way forward.²⁵

[Figure 4 near here]

However, our archival research has identified another way forward, although it does not meet the specific needs of the Environment Agency. We located the original color separations used to prepare printing plates for the southern national summary sheet, as shown in figure 4 and covering most of England and Wales, and about twenty of the one inch sheets

published by the Stamp survey. As each separation defines just one of the color layers, they are far more easily and accurately vectorized. Our current research, following our own agenda with funding from the Frederick Soddy Trust, is constructing vector versions of this national summary sheet plus the detailed surveys of Birmingham and Dartmoor, enabling detailed studies of change on the urban and moorland fringes. “Modern” land use is recorded via the Land Cover 2000 digital mapping created from satellite data by the Centre for Environment and Hydrology. One challenge is distinguishing differences due to historical change from differences due to the very different survey methods, but we also have access to a detailed survey of the Brighton area made in 1996-7 by the Land Use UK project, using the same methods as the LUSGB but at a very similar date to the Land Cover data.²⁶

If one conclusion is the need for historical GIS to explore the potential of raster-based GIS techniques, another is the need to balance contract research with our own agenda. Our work on long-run land use change, which also includes analysis of farm census data, would have been impossible without the support of DEFRA and its agencies; but it would also have been impossible without the goodwill of numerous map librarians, of the LUSGB copyright holder and of the Frederick Soddy Trust, all gained through our commitment to creating a public resource.

Property Services

“Old maps” are of course essential to defining property boundaries and to understanding the past history of properties, and academic historical geographers have sometimes acted as consultants in legal disputes.²⁷ However, individual disputes are unlikely to justify construction of a GIS, and the two examples discussed here use specialized historical GISs to drive web sites that serve mass markets; not lawyers involved in disputes but lawyers handling routine property transactions. Anyone purchasing a property, whether a large industrial site or an ordinary house, needs information on associated liabilities. Some

will be obvious, such as council or property taxes. Others will relate to the site's past history and be far less obvious.

Firstly, the Landmark Information Group are a commercial partner of the Ordnance Survey and have created what is essentially a very large GIS built primarily from nearly a million historical maps at six inch-to-one mile (1:10,560) and more detailed scales for Britain between c. 1848 and c. 1995;²⁸ by contrast, the largest non-commercial historical maps database is that constructed by the US Geological Survey's Historical Quadrangle Scanning Project, covering approximately 200,000 maps.²⁹ Landmark's database then forms the basis for a number of distinct products sold on-line. These include the old-maps.co.uk web site, which straightforwardly sells reproductions of historic maps to the general public. However, this is a sideline compared to Envirocheck:

The Envirocheck Historical Data Report identifies historical land use, such as: potentially contaminative industrial uses; potentially infilled land; historical tanks and energy facilities; data captured from Landmark's unique set of historical building plans such as potentially contaminative features from historical building plans, which includes asbestos and areas cleared due to enemy action (bomb damage); list of the historical building plans analysed for your selected area. This report comprises of an A4 datasheet, an A3 mapsheet at 1:10,000 scale and up to 16 mapsheets at 1:2,500 scale. A separate report is produced for each site slice.³⁰

The second example directly involves the GB Historical GIS, and has been a significant source of income to cover the running costs of our web site, *A Vision of Britain through Time*. Under a series of laws starting with the Tithe Commutation Act of 1836, owners of "rectorial land" in England and Wales are liable for the cost of repairs to the chancels of parish churches.³¹ In recent years, a small number of property owners have found themselves suddenly liable for very substantial sums, the largest ever claim being £186,969. Establishing

exactly which properties are liable is very time consuming, but only properties located within certain “tithe districts”, as defined in 1836, are potentially liable, amounting to about 35 percent of all parishes;³² and those property owners form a fairly lucrative market for specialized insurance.

Curiously, the Great Britain Historical GIS project has been involved in the supply of two quite different maps of nineteenth century parishes to two different companies, both serving this market. Firstly, research by Roger Kain and Richard Oliver of Exeter University was based mainly on local surveys from the late eighteenth and early nineteenth century, enclosure and tithe maps, from the right period but unavoidably leaving some parts of the country uncovered. The Exeter project created a “digital map” which in fact consisted of a series of separate Adobe Acrobat files, one for each New Popular edition one inch map, consisting of vector boundary lines drawn on top of scans of the one inch maps. The GB Historical GIS created a true GIS from these by converting the vector data to ArcGIS, constructing a true polygon topology, geo-referencing the data and assembling the data from all the maps into a single national coverage. These data have been licensed to Conveyancing Liability Solutions Ltd (<http://www.clsi.co.uk>).³³

Our own research used a different methodology, starting by digitizing the Civil Parish system as it existed in the late 1900s, using comprehensive and unproblematic Ordnance Survey maps. We then assembled all the textual descriptions of parish boundary changes in published census reports, including the very extensive changes resulting from the Divided Parishes Acts of 1876 and 1882, as listed by the 1891 census. Unraveling these changes to map the pre-1876 parish geography then involved a project member spending two years in the National Archives, using a range of cartographic sources.³⁴ The resulting vector boundaries, with some further enhancements, have been licensed to Pinpoint Chancel Search Ltd (<http://www.pinpointchancel.co.uk>). Creating a commercial product involved our linking

the boundaries to a digitized version of another source in the National Archives, the Record of Ascertainments (class IR 104) which identifies the actual liabilities. This linkage involved extensive use of our historical gazetteers.

Chancel repair liability is financially important to the GB Historical GIS but is obviously a somewhat special case, especially as from 2013 the liability will exist only if it forms part of the registered title, and the market for liability searches will presumably disappear; fortunately advertising income from our web site is now sufficient to cover its running costs. However, the system built by Landmark Information clearly dwarves anything built by an academic historical GIS project. The challenge is to unlock the analytic potential such systems clearly contain. While the government agencies discussed in the previous section actively supported our publishing the raster LUSGB data, licensing data to commercial companies obviously stops it being made available to all-comers, and has arguably inhibited academic use. Issues of “commercial confidence” have certainly inhibited the account given here, so for example nothing has been said about the two points where we needed to formally respond to hints of possible legal action against us.

Conclusions

This chapter has hopefully shown both how GIS techniques can be applied to a range of broadly historical challenges outside the accepted realm of “historical GIS”, and how our approach has had to adapt to meet the needs of workers in other fields. One general comment is that although many of our collaborators have had limited actual knowledge of the practicalities of GIS, there is a high level of general awareness of GIS, and enthusiasm for its potential.

One final large question needs to be answered: why bother with the kinds of activity described here. One obvious answer is to make money, and this is of course why commercial companies such as Landmark Information have become involved in historical GIS. However,

this has clearly not been the main justification for the GB Historical GIS: our main source of funding remains, by far, grants not contract work or other commercial earnings.

Firstly, the construction of a national historical GIS is one of the most expensive possible pieces of academic infrastructure construction in the humanities and social sciences. It is almost impossible to justify the cost, even to grant giving bodies, if the only benefit will be to academic historians. Note that the other beneficiaries may still be academic researchers, as with the GBH GIS's collaborations with medical researchers. However, this need to demonstrate a wider public benefit is particularly strong when the funding source is a digitization program. That said, we would always emphasize that while the *Vision of Britain* web interface is of limited use to GIS researchers it clearly serves a large audience of academic historians researching particular localities.

Secondly, although our non-grant income is a small percentage of the total, it can be spent more flexibly and for purposes for which grants are not available. Expenditure from grants is always carefully monitored. Many grant-giving bodies pay only in arrears, based on evidence of actual expenditure. It is therefore impossible to build up any reserves, and consequently to keep staff in post during the almost inevitable periods when grant applications go badly. Conversely, a major reason why the GBH GIS has been able to keep staff in post over long periods despite lacking any core funding is reserves built up from contract work, although universities of course vary in how far such money is available to the researchers who brought it in. Another factor in sustaining staff has been developing relationships with small charitable trusts, whose grants have also been a small percentage of the total but very important in hard times. "Commercial" earnings are also necessary to sustain the *Vision of Britain* web site, as there are *no* sources of grant funding for this kind of running cost.

Thirdly, an increasing factor in the UK is the research assessment framework managed by the higher education funding councils. Up to now, this has been based almost entirely on academic peer review, and consequently has emphasized traditional publications, especially in peer reviewed journals. However, under the new Research Excellence Framework, which is replacing the Research Assessment Exercise, 20 percent of funding will be allocated based on “economic and social impact”; and the work of the GB Historical GIS described here very clearly does provide just such a wider economic and social benefit.³⁵

However, writing more personally, I would argue that the strongest reason for a historical GIS researcher becoming involved in these wider areas is simply that it is more interesting and fulfilling to work with many distinct communities, learning about their different perspectives and finding common ground. As it happens, these particular words are being written on a train to Scotland, for a HALCyon meeting at the Royal College of Physicians in Edinburgh, two days after a day spent with environmental officials and commercial GIS consultants focused on land use change; but our contribution is always about working with historical sources using GIS tools. This is vastly more fun than trying to convince cultural historians and cultural geographers of the “academic value” of our work, even if you do not care about delivering “economic and social benefits” to the wider society we live in; and I do, a great deal.

Notes

¹ H. Southall, “Rebuilding the Great Britain Historical GIS, part 1: Building an indefinitely scalable statistical database”, *Historical Methods* 44 (2011), 149-59; H. Southall, “Rebuilding the Great Britain Historical GIS, part 2: A geo-spatial ontology of administrative units”, *Historical Methods* 45 (2012), 119-34. The third part will cover the inclusion of more qualitative content, and the design and use of the web site.

² D. Barker, *Fetal and infant origins of adult disease* (London: British Medical Journal Publishing Group, 1992); L. Berney, D. Blane, G. Davey-Smith, P. Holland, P., “Life course influences on health in early old age”, in H. Graham, ed., *Understanding Health Inequalities* (Milton Keynes, Open University Press, 2001), 79-95.

³ Y. Ben-Shlomo and G. Davey Smith, “Deprivation in infancy or in adult life: which is more important for mortality risk?” *Lancet* 337 (1991), 530-4.

⁴ Department for Education and Skills *Towards Understanding Sure Start Local Programmes: Summary of Findings of the National Evaluation* (Nottingham: Department for Education and Skills, 2004). <http://publications.education.gov.uk/eOrderingDownload/NESS-SF-2004-007.pdf> [22 July 2012].

⁵ R.M. Martin, D. Gunnell, J. Pemberton, S. Frankel and G. Davey Smith, “Cohort Profile: The Boyd Orr cohort—an historical cohort study based on the 65 year follow-up of the Carnegie Survey of Diet and Health (1937–39)”, *International Journal of Epidemiology*, 34, (2005), 742-9.

⁶ For an introduction to the LS, see L. Hattersley, R. Creeser, *Longitudinal Study 1971–1991: History, organization and quality of data*, Longitudinal Study Series no. 7 (London: HMSO, 1995).

⁷ D.P. Strachan, D.A. Leon, and B. Dodgeon, “Mortality from cardiovascular disease among interregional migrants in England and Wales”, *British Medical Journal*, 310 (1995), 423-7.

⁸ L. Simpson, “Geography conversion tables: a framework for conversion of data between geographical units”, *International Journal of Population Geography*, 8 (2002), 69-82

⁹ S. Curtis, H.R. Southall, P. Congdon and B. Dodgeon, “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*, 58 (2004), 57–74.

¹⁰ M. Wadsworth, D. Kuh, M. Richards and R. Hardy, “Cohort Profile: The 1946 National Birth Cohort (MRC National Survey of Health and Development)”, *International Journal of Epidemiology*, 35 (2006), 49–54.

¹¹ E. Garrett, A. Reid, K. Schürer and S. Szreter, *Changing Family Size in England and Wales: Place, Class and Demography, 1891–1911* (New York: Cambridge University Press, 2001).

¹² E.T. Murray, H.R. Southall, P. Aucott, K. Tilling, D. Kuh, R. Hardy, Y. Ben-Shlomo , “Challenges in examining area effects across the life course on physical capability in mid-life: Findings from the 1946 British birth cohort”, *Health and Place* 18 (2012) 366–74.

¹³ For example: S. Szreter, “The importance of social intervention in Britain’s mortality decline c.1850–1914”, *Social History of Medicine*, 1 (1988), 1-38; M.R. Haines, “Socio-economic Differentials in Infant and Child Mortality during Mortality Decline: England and Wales, 1890-1911”, *Population Studies*, 49 (1995), 297-315; I.N. Gregory, “Different places, different stories: Infant mortality decline in England and Wales, 1851-1911” *Annals of the Association of American Geographers*, 98 (2008), 773-94.

¹⁴ P.J. Aucott, A. von Lünen and H.R. Southall, “Exposing the history of Europe: the creation of a structure to enable time-spatial searching of historical resources within a European framework”, *OCLC Systems and Services* 25 (2009), 270-86.

¹⁵ The author is a member of this working group. Further details of the survey can be obtained from Patricia McGuire of King’s College, Cambridge (archivist@kings.cam.ac.uk).

¹⁶ Until 2002, this inspection function was carried out by the National Register of Archives, who then merged with the Public Record Office to form the National Archives; and it was the NRA whose support was crucial to obtaining UK National Lottery funding.

¹⁷ J. Stevenson, “Place names: we would be lost without them”, *Archives Hub Blog* (7 Dec 2009). <http://archiveshub.ac.uk/blog/2009/12/place-names-we-would-be-lost-without-them> [22 July 2012]; H.R. Southall, “Alternative ways of indexing by geography”, *ARC magazine* 254 (2010), 17-19.

¹⁸ <http://www.nationalarchives.gov.uk/documents/information-management/naming-rules.pdf> [22 July 2012]. For Scotland we used the SCAN gazetteer of Scottish places, which had been created since the NCA report, but this was anyway partly based on the 1971 census listing.

¹⁹ A more extensive account of this architecture is provided by Southall, “Rebuilding the Great Britain Historical GIS, part 2”.

²⁰ <http://dublincore.org> [22 July 2012].

²¹ <http://eac.staatsbibliothek-berlin.de> [22 July 2012].

²² <http://www.opengeospatial.org> [22 July 2012].

²³ http://www.naturalengland.org.uk/about_us/whatwedo [22 July 2012].

²⁴ Andrew Baker, personal communication, November 2009.

²⁵ B. Baily, M. Riley, P. Aucott and H. Southall, “Extracting digital data from the First Land Utilisation Survey of Great Britain: Methods, issues and potential”, *Applied Geography* 31 (2011) 959-68. One published study making some use of these data, for the Peak District, is M. Dallimer, D. Tinch, S. Acs, N. Hanley, H.R. Southall, K.J. Gaston and P.R. Armsworth, “100 years of change: examining agricultural trends, habitat change and stakeholder perceptions through the 20th century”, *Journal of Applied Ecology* 46 (2009), 334–43.

²⁶ R. Walford, ed., *Land Use-UK: A Survey for the 21st Century* (Sheffield: Geographical Association, 1997).

²⁷ For some examples see R. Janiskee, “Socially and ecologically responsible historical geography”, *Environmental Review*, 4 (1980), 35-40.

²⁸ <http://www.landmark.co.uk/page/historic-data> [22 July 2012].

²⁹ G.J. Allord and W.J. Carswell, “Scanning and georeferencing historical USGS quadrangles” (Madison: US Geological Survey, 2011).

http://pubs.usgs.gov/fs/2011/3009/fs20113009_sept2011.pdf [22 July 2012].

³⁰ http://www.envirocheck.co.uk/envirocheck/historical_land_use_report.jsp [22 July 2012].

³¹For much more about chancel liability and the relevant historical records, see National Archives, “Chancel repair liabilities in England and Wales” (Legal Records Information Leaflet 33): <http://www.nationalarchives.gov.uk/records/research-guides/chancel-repairs.htm>, [22 July 2012].

³² <http://www.csl.co.uk/contentmanagement/Page.aspx?PageID=254&b2cmode=False> [22 July 2012].

³³The original Kain and Oliver data are held by the UK Data Archive as SN 4348: “Historic Parishes of England and Wales: an Electronic Map of Boundaries before 1850 with a Gazetteer and Metadata”. The GIS constructed by the GBH GIS is similarly held as SN 4828: “GIS of the Ancient Parishes of England and Wales, 1500-1850”.

³⁴ I.N. Gregory, C. Bennett, V.L. Gilham and H.R. Southall, “The Great Britain Historical GIS: From maps to changing human geography”, *Cartographic Journal*, 39 (2002), 37-49.

³⁵ <http://www.ref.ac.uk> [22 July 2012].

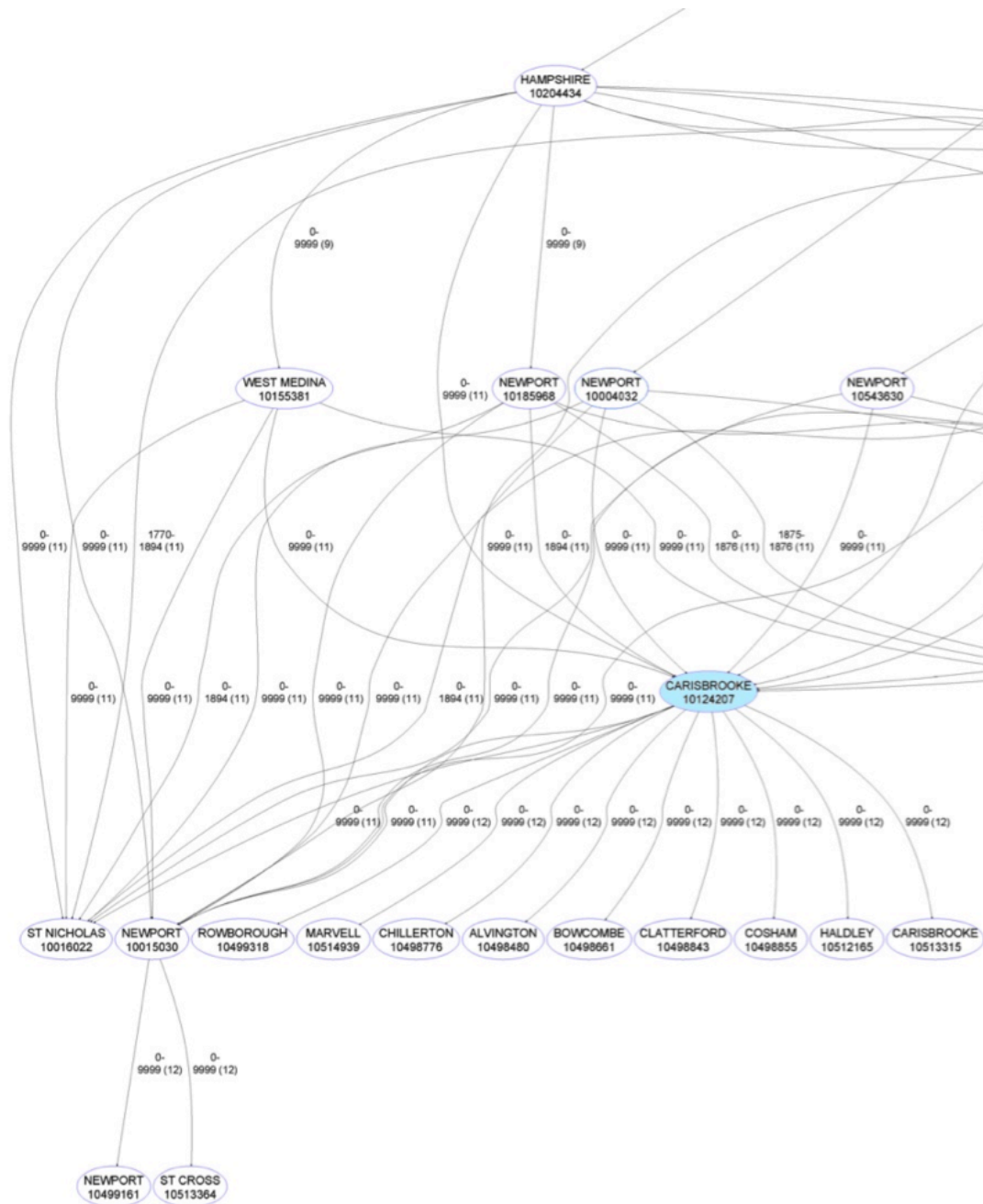


Figure 1: The Parish of Carisbrooke, Isle of Wight, and its “IsPartOf” relationships. The units “below” Carisbrooke are mostly manors, added from the UK National Archives’ Manorial Documents Register, but include the chapelries of St Nicholas in the Castle and Newport. By the nineteenth century the market town of Newport was more important than the fortified village of Carisbrooke, so Carisbrooke came under the Urban Sanitary District of Newport, the Municipal Borough of Newport and so on. All these units were within the Ancient County of Hampshire, but not the later Administrative County of Hampshire as the Isle of Wight was a separate Administrative County. Visualization created by Vojtech Kupca of the University of Umeå.

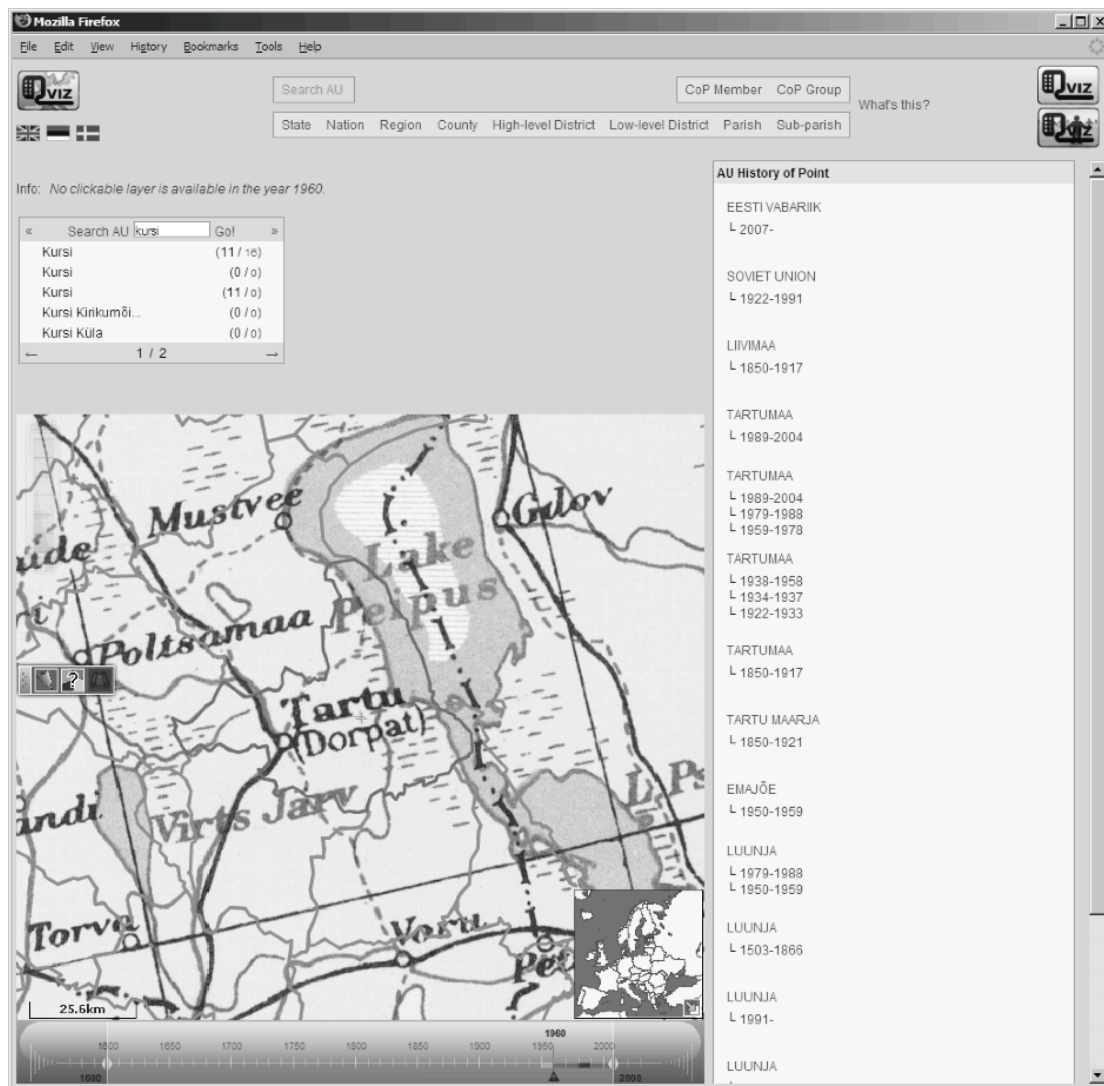


Figure 2: The QVIZ User Interface presenting information about a location in Estonia and the administrative units which covered it at different dates. The location is in fact the city of Tartu, the home of the Estonian National Archives. Just as with Carisbrooke, but unlike many locations in the United States, there has been a complex sequence of historical units. The earliest unit listed here, “Luunja”, covered Tartu from 1503 to 1866. This interface was created by the QVIZ teams at AS Regio (Estonia) and the University of Umeå, accessing a data structure created by the GBH GIS team.

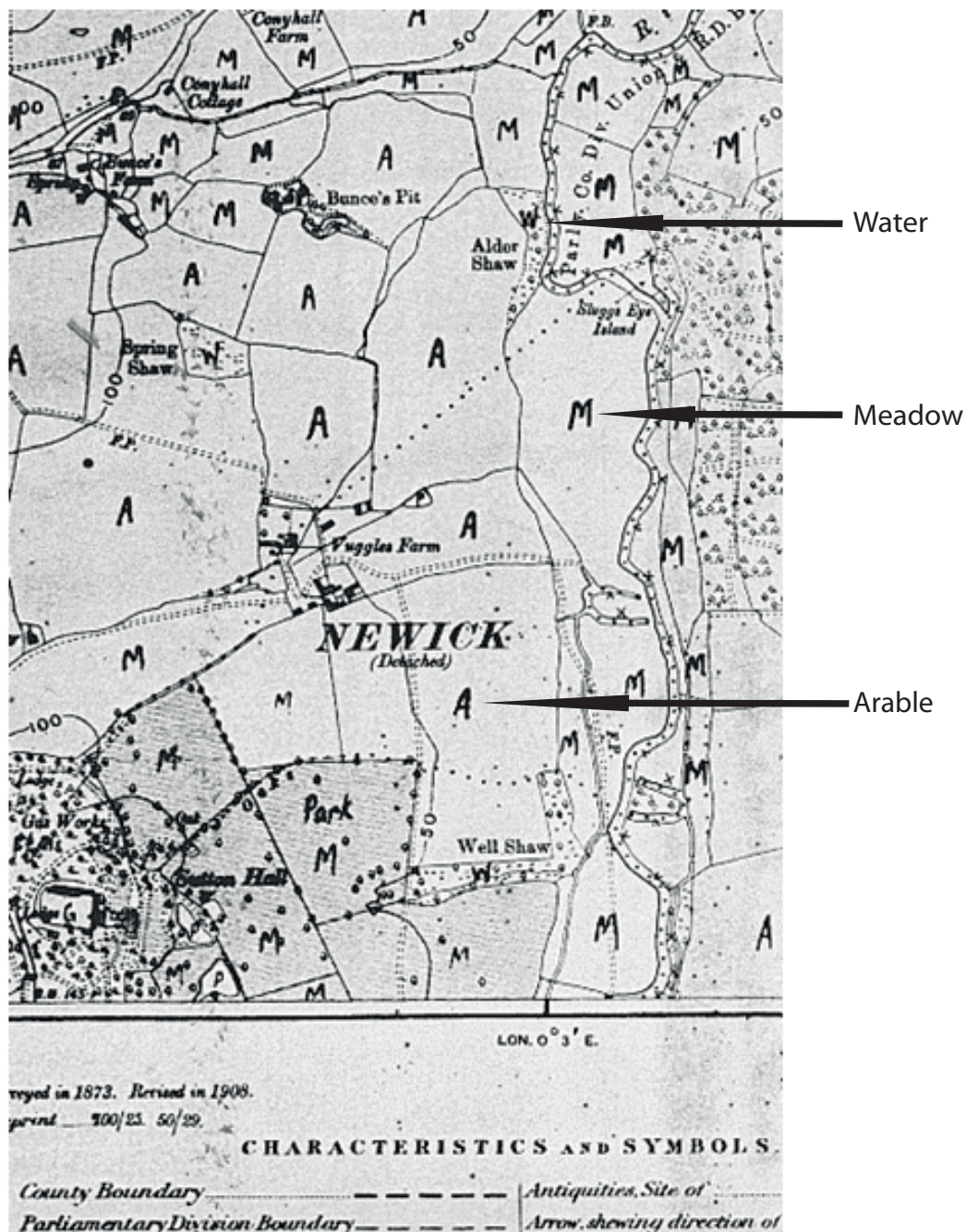


Figure 3: Sample Field Survey Sheet from the Land Utilisation Survey of Great Britain, showing part of Newick, Sussex. These six inch to one mile (1:10,560) Ordnance Survey maps were issued by the Survey to schools, and then by the schools to individual pupils who hand-annotated them to record land use: a very simple set of polygons and attributes. They were then returned to the Survey in London and, although some have been lost, the London School of Economics Archives hold a large collection. Reproduced by kind permission of Audrey N. Clark and Giles Clark.



Figure 4: Land Utilisation Survey of Great Britain National Summary Sheet (South): Color separation for the green (pasture) layer. The Survey team in London collated the information from the Field Survey Sheets to create a set of maps at one inch to one mile scale (1:63,360), most of which have been published. They then further condensed the information to create two national summary sheets at ten miles to one inch scale (1:625,000). The published maps are difficult to process due to the mix of colors, and overprinted names and topographic features. Color separations such as this were created as a stage in the printing process, and contain just one of the color layers making image processing much simpler. Ideally, all the separations would have been preserved but the Survey had no obvious reason for doing this, so we are fortunate that about ten per cent of the separations were preserved by a member of the Survey's staff, and passed to the London School of Economics Archives by a relative. Reproduced by kind permission of Audrey N. Clark and Giles Clark.