

Great Britain Historical Gazetteer/GIS

Humphrey Southall (University of Portsmouth)

Introduction

Possibly the single most important commitment of the Great Britain Historical GIS project under its funding from the New Opportunities Fund is the construction of a systematic historical gazetteer covering the administrative units of Great Britain over the last two centuries. This is important both as a reference resource in its own right and to tie together our diverse but always place-specific content: we promised “Database-driven ‘home pages’ for **every** administrative unit – about 25,000 pages, for just about every town and village”, and the entries in our administrative unit gazetteer define which ‘places’ we will have home pages for.

This document describes a growing resource. Our aim is to make information on the parishes of Great Britain available to the general public by the end of 2003, from a web site operated by our partners at EDINA. During 2003, two more limited on-line systems are being made available to interested people from our development server in Portsmouth:

- **Demonstration system:** This interface is deliberately limited to content for the Isle of Wight, combining information from Youngs’ *Local Administrative Units*, the National Register of Archives’ *Manorial Documents Register*, two descriptive gazetteers published in the late nineteenth century, statistics from the Census of Population and, crucially, polygons describing the changing boundaries of parishes and districts. The Isle of Wight was chosen as being an area covered by the English system of civil administration with enough internal complexity to be a worthwhile test, but where boundary changes would never involve units in another county. The demonstration system is currently mainly textual, but will evolve towards the much more graphical interface planned for use by the general public. It can be accessed at:

http://www.gbhgis.org/demo_gaz

This address is probably only valid for the current year, and should therefore not be linked to from other web sites, but can be circulated.

- **Evaluation system:** The database the Demonstration system accesses already goes substantially beyond the Isle of Wight, including counties for the whole British Isles and parishes for Scotland. We expect to add parishes for four additional English counties during February 2003: Hampshire, Herefordshire, Lancashire and Norfolk, plus London, these being chosen to cover a range of areas and so test the system’s ability to include all variants. We do not want what is primarily a development system to go into widespread use, but we do want people with the knowledge to check our work to be able to access it. We are therefore creating a second interface which accesses the entire database but focuses on the more formal information on names and hierarchical

relationships, as found in Youngs. This system will evolve towards the alternative 'expert search' interface we are planning for the final system, and we need feedback on what this should contain. Information about this second interface will be circulated to people who join our new mailing list, gbhgis@JISCmail.ac.uk, details of which are accessible from the above web site.

Overview:

Our computer system is designed to be many things to many people, which makes it harder to explain than demonstrate. It already holds quite diverse content, including textual descriptions, statistics and digital boundaries. As it already includes, for example, all the information needed to generate a sequence of parish-level population density maps for the Isle of Wight between 1801 and 1971, it has a real claim to be a Geographical Information System. However, the software to permit it to create such maps is still being written and for now it looks more like a gazetteer, which is how this document discusses it. The system is in fact two tightly linked but distinct gazetteers, one specifically concerned with administrative units, the other containing computerised descriptions of places drawn from gazetteers published in the 19th century. The 'Descriptive Gazetteer' is further discussed below.

As an administrative units gazetteer, its clearest precursor is Youngs' *Local Administrative Units*, and that is also the largest single source for the core gazetteer. However, it should be seen as a new resource, not a 'computerised Youngs', for a number of reasons:

- **It is a (spatial) database, not a book:** It would have been possible to simply computerise the text of Youngs and place it on the web, but this would not have served our wider aims. Instead, we have placed information from Youngs into a very highly structured database, which in turn supports a web site in which all references to other related units appear as hyperlinks, taking users to the entry for that other unit. This is useful in itself, but also means the system has been checked for consistency in ways impossible in a book; in the process, we have found many inconsistencies between Youngs' lists of parishes in each district and the districts named in his parish entries. Further, the system is a **spatial** database: it records parishes, districts and counties not simply as names in a list or even as points on a map but as sequences of polygons based on changing historic boundaries; this in turn lets us show them on maps, and provide a unique facility enabling users to look up units starting with a postcode. There has been some price to pay in greater rigidity, but we hope it is justified by the greater consistency and functionality. NB there are many places in our system where unstructured information can be added as notes.
- **It is a systematic record of corporate bodies:** Like Youngs but unlike most gazetteers, our main gazetteer is a record of administrative units, not 'places' or 'place-names'; one aspect is that we plan to support the International Council on Archives ISAAR (CPF) standard (International Standard Archival Authority Record for Corporate Bodies, Persons and Families), as well as the Alexandria Digital Libraries proposed standard for Gazetteer Content

(standards support is not discussed in detail in this edition of this document, but our approach is compatible). ‘Place-names’ are nicknames for vaguely defined geographical areas, but administrative units are created, abolished, named and altered by clear legal processes. While feature types thesauri are a major focus of current discussion of digital gazetteers, our feature types are generally clearly specified in our sources. One example of how our focus leads us to take a slightly different approach from Youngs is in our treatment of English local government districts: an Urban District which was ‘promoted’ to a Municipal Borough has two separate entries in Youngs, but we give it only one as it was clearly the same body. NB while this approach is very helpful with districts, counties and regions, problems arise at parish-level and below, where the administrative machinery associated with each unit was modest and informal, and at the national and semi-international level, where most bodies have unique aspects.

- **It is closely linked to digital boundary mapping:** The system described here may look very different from the Geographical Information System this project previously built, but it is designed to include all the boundary mapping we have created: our own earlier work on the changing Civil Parishes of England and Wales since the 1870s; a separate record of Ancient Parishes created by Roger Kain and Richard Oliver of Exeter University, and converted into a GIS by us; and a new GIS we have built for Scotland. As discussed below, these data support quite new functionality including postcode-based searching, but their inclusion has influenced the system’s design. In particular, we generally combine different kinds of units into a single *Unit Type*, kinds being distinguished by status values, so that a map of all the units of a given type at a given date covers the whole country. One specific consequence is that where two sets of boundaries, and therefore populations, existed simultaneously for what some would regard as the same unit we require that two distinct units exist. This means that Ancient, Registration and Administrative Counties of the same name are quite distinct units, and that Ecclesiastical Parishes are kept distinct from Ancient/Civil Parishes of the same name and general location.
- **It must provide systematic access to a very large body of census and vital registration data:** Our GIS was originally built not simply as a record of changing boundaries for use by local historians, but as a tool for systematically structuring a very large body of statistics derived from historical census reports, vital registration data and similar; and to make it possible to analyse long-term demographic change despite drastic changes in the reporting units. All this statistical information will form part of our final web site and be accessible via the gazetteer. The most obvious practical consequence is for the units we have to cover: we are including Registration Districts and sub-Districts, not listed by Youngs, while we give relatively low priority to ecclesiastical geographies. This said, our architecture is very flexible and can include almost any conceivable administrative geography.

Having emphasised the various ways in which we are not exactly following Youngs’ approach, one goal should be clearly stated: we aim to include all the unit names

listed in Youngs, and equivalent sources for other parts of Britain, and give these names primacy over the many variant names that appear in our census transcriptions.

Gazetteer Sources

Our aim is to create an authoritative reference resource, and acceptance will be much easier to achieve if the system is based on established authorities, conventionally published. More specifically, we aim to follow as closely as possible the authorities identified in the National Council on Archives' *Rules for the Construction of Place Names* (see <http://www.hmc.gov.uk/nca/rules3.htm>), and the inevitable copyright issues have largely been resolved.

England:

Youngs' *Local Administrative Units* is the established authority, and we have had permission from the Royal Historical Society to computerise it for some time. However, given that our goal is to construct a database which supports a web site in which, for example, all cross-references are hyperlinks, Youngs is an extremely complex resource to computerise. We had a good look at building a database from it, but the result would be a very awkward system for no very good reason: Youngs' structure seems to be mainly the result of a need to achieve extreme compression so the books could be economically published, which is not an issue with a web site. We then looked at creating an equivalent resource from other primary sources. Our final solution is a hybrid, which uses selected information from Youngs as the core but adds a great deal of information drawn from census reports.

Two substantial changes to Youngs' organisation of units, and one 'clarification', have proved necessary:

- **Parishes:** Youngs treats most parishes as single units combining Ancient, Civil and Ecclesiastical roles. While the overall gazetteer architecture allows a single unit to have several status values, consecutively or concurrently, Ecclesiastical Parishes often had quite different boundaries and therefore populations from their civil namesakes. In our gazetteer, therefore, a single parish-level unit may have both 'Ancient Parish' and 'Civil Parish' status, the transition from one to the other not being precisely dated, any Ecclesiastical Parish of the same name will exist as a separate entry, generally linked to the earlier AP via a 'SucceededBy' relationship.
- **Boroughs:** Youngs organises district-level units for civil administration into Hundreds, Boroughs and Local Government Districts, Boroughs covering *both* ancient boroughs and post-1894 Municipal and County Boroughs. Partly to simplify working with statistical tables, we have included ancient boroughs with Hundreds, as 'Ancient Districts', while Municipal and County Boroughs are included with Urban and Rural Districts.
- **Counties:** Youngs lists parishes by Ancient Counties, but also identifies distinct Poor Law Counties and Administrative Counties. The lists of Ancient and Poor Law counties are identical, but their boundaries differed very

substantially. Administrative County boundaries more closely resembled those of Ancient Counties, but sub-divided several of the latter; for example, East and West Sussex and the three Parts of Lincolnshire. Our system defines three distinct types, ANC_CNTY, PR_CNTY and ADM_CNTY, and links the units of different types but the same name via SucceededBy relationships (and, slightly less formally, by their polygons overlapping as well as their names matching). This distinction is also maintained for Wales but not for Scotland or Ireland.

We are also including additional types of unit not covered by Youngs:

- **Registration Districts:** These were largely identical to Poor Law Unions, and at present we are trying to use a single unit type, PR_DIST, and record the differences via status values, PLU and RegD, which can have separate dates. Most PR_DIST units will have both status values.
- **Registration Sub-Districts:** Our statistical database includes much demographic data for these so we must construct our own authority list. The method currently proposed is to take an initial set of names from the 1911 census parish table; standardise punctuation; standardise place-names to those given by Youngs for associated Poor Law Unions and/or parishes; then add earlier units using the lists of changes in the Registrar General's *Annual Reports* and *Decennial Supplements*.
- **Manors:** For a limited number of counties, we can include information from the National Register of Archives' *Manorial Documents Register*. This includes preferred and alternate names for manors, and the names of the ancient parishes that contained them, but no information on chronology or change.

Wales:

The NCA Rules specify Melville Richards' *Welsh Administrative and Territorial Units*, published for the Board of Celtic Studies by the University of Wales Press in 1969. We are in discussion with the University of Wales. The information in Melville Richards is simpler than that in Youngs, but will fit into the same framework. Melville Richards includes no information on Poor Law Unions so this will need to be constructed based on information in the census reports.

Scotland:

The Scottish Archives Network (SCAN) is providing us with the authority lists they have constructed for **counties**, **parishes** and **burghs**. Their counties and parishes have already been added. Burghs are not yet complete, but must be finalised by the end of their own funding in March 2003. Their lists include basic information on dates of creation and abolition, and on hierarchical relationships.

NB:

- There are many other Scottish administrative units that the census made some use of, but they are not being covered by SCAN and we have no current plans

to research them ourselves. Some work on the ‘districts’ used by the C20 census may be unavoidable.

- The status of Scottish burghs is complex. The census reports record that the boundaries of a given burgh at a given date could vary slightly between those of, for example, the municipal burgh and the police burgh. However, we have no systematic sources of information on those differences in areas and therefore population, so we will be ignoring them. We can, however, record the different burghal statuses via the gazetteer’s ‘status’ table.
- SCAN do not claim any specific authority for the spellings of the names in their gazetteer. The National Council on Archives ‘Rules’ specify the Index of Place Names from the 1971 census reports as the name authority for Scottish administrative units. The General Register Office have confirmed that our permission to digitise the census reports covers this list (their original letter specified ‘statistics’ from the reports, so confirmation was needed).

Other:

Our commitment is to create a single resource covering the whole of Great Britain, and our long term aim is to cover the whole British Isles (without necessarily using that term; Norman Davies’ book *The Isles* is a useful guide in this area). This raises some issues which would not arise in a narrower creation of specifically English, Welsh and Scottish gazetteers:

- **States and Nations:** Youngs does not cover ‘England’, and it is not entirely clear that England currently has any specific administrative existence. Somewhat arbitrarily, the current gazetteer has a top level called ‘World’, within which are three ‘States’: the United Kingdom, the Irish Republic and the Isle of Man; the United Kingdom in turn contains four ‘Nations’: England, Wales and Scotland plus Ireland to 1922 then Northern Ireland. Obviously, whole books can be and have been written about the precise legal status of these entities.
- **Ireland:** Our funded commitments specifically exclude Ireland. However, a basic list of Irish counties, taken from the NCA Rules, has been added to enable descriptive material from Bartholomew’s *Gazetteer* to be integrated in a uniform way.
- **Isle of Man:** We made no specific commitment but we may be able to include it within current funding, unlike Ireland. We are in contact with the Isle of Man Record Office.
- **Channel Islands:** Like the Isle of Man, these are not part of the United Kingdom but do, for example, use British postcodes and may be included if time permits. We are in contact with the Island Archives Service in Guernsey.
- **Special Cases:** Within the last 200 years, the external boundaries of the UK have been stable other than in Ireland, and we already record which counties became part of the Republic. The boundary between Wales and England

involves some complex changes and relationships, including parishes which were simultaneously in an English administrative county and a Welsh registration county. However, these should not pose major problems as we are using the same typology both sides of the border. Berwick on Tweed has moved between England and Scotland, which may pose a real problem but is a single case. Taking the system back into earlier periods would involve additional states and, for example, the acquisition of the Shetland Isles from Denmark.

Boundary changes:

Much of the content of the parish entries in Youngs is details of boundary changes. Melville Richards and the SCAN gazetteer lack this information. However, most of Youngs information covers the late 19th and the 20th centuries, and the project already has very extensive transcriptions of fuller information on the same changes that were published in census and vital registration reports for England and Wales.

Unfortunately, the format of these boundary change lists varies greatly, but the prototype system described below already contains data from the 1891 census, covering the two Divided Parishes Acts, and the 1930s, covering the period of the county reviews. Similar information was published for Scotland in census and RG reports for 1921 onwards, while the extensive changes in 1891 are documented elsewhere.

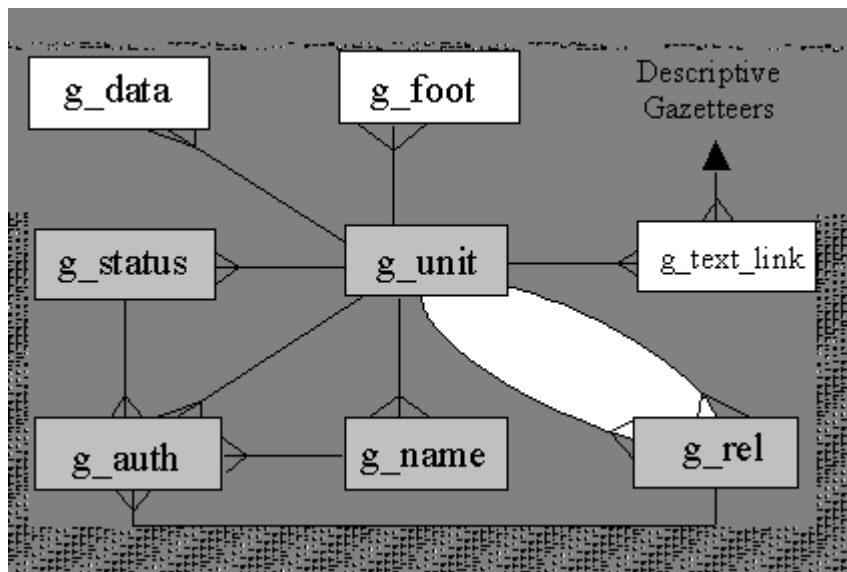
Variant names:

Youngs provides some information on variant names. Melville Richards provides alternative Welsh and English names for some places, and the prototype system is multi-lingual. We are marking those names given in our authorities as preferred, but the system can hold an infinite number of alternative names for each unit. We are following Youngs in capitalising all names and removing almost all punctuation, and different approaches to capitalisation and punctuation do not create alternate names. However, variations in spelling do. For our gazetteer to work as an access route to our own statistical holdings, we must treat all names used by the census, other than obvious printing errors, as valid alternative names; the system holds information on the actual source of each of these names.

Data Model

There is a great temptation to try to model the complex hierarchies of administrative units in the actual structure of the gazetteer, and as various special cases emerge the structure gets steadily more complex. Our system handles the complex inter-relationships between units via a simple but flexible structure. This structure also enables each unit to change its name, hierarchical relationships and location. The basic structure involves five tables, although several additional tables define the permissible values for types of unit, the changing statuses of units, the relationships between units and so on (the contents of these tables is essentially the information listed here in the appendices). These five tables are in grey in the following diagram;

the other three hold statistical data (**g_data**) or polygonal footprints (**g_foot**), or provide linkage to our descriptive gazetteer (**g_text_link**), and are discussed later: NB this is the actual structure of the database driving the prototype web site:



The **g_unit** table defines which units existed, but holds very limited information: an ID number generated by the system, a **unit_type** value which is a fixed attribute, dates of creation and abolition, and a notes column (NB every record in every table can hold notes).

- *All* names of units, preferred and variant, are held in **g_name**.
- The **g_status** table holds characteristics of units which may change over time, or be held simultaneously: it will record, for example, whether a local government district was an Urban District, Municipal Borough or County Borough, or whether an earlier district was, most commonly, both a Poor Law Union and a Registration District, or just one or the other (in these two examples, the *g_unit_type* values are, respectively, LG_DIST and PR_DIST).
- The **g_rel** table holds a range of relationships between units, including:

IsPartOf	Basic hierarchical relationships (but NB a unit can be part of more than one higher level unit simultaneously).
WasSucceededBy	This handles a wide range of relationships, generally between units of different types, and is a kind of 'see also'; for example, it is used to relate Ancient, Registration to both Registration Counties and Administrative Counties, and Registration Counties to Administrative Counties.
ReducedToEnlarge ReducedToCreate AbolishedToEnlarge AbolishedToCreate	These variants all record boundary changes, and the table includes fields to hold information about the area transferred, such as area, number of houses and number of people.

BoundaryChange	Unlike the above, these entries lack information about the other unit involved, which is unavoidable where our source is Youngs as he names the other unit only where one of the units was either abolished or created.
Event	The inclusion of this type enables the g_rel table to hold any other event affecting a unit.

- The **g_auth** table is used to hold information on gazetteer sources, and exists to enable multiple authorities to be associated with any item for example, the place within the census reports we transcribed a boundary change from and the Ministry of Health Order that the census cites; there are a few cases where multiple statutory orders are cited for a single change.

Assessment

Although this structure is far more elegant than earlier models in which each type of units had its own table, and is more or less infinitely extensible, we had a number of concerns:

- Could data be loaded into it and then extracted? The initial ‘Isle of Wight’ gazetteer obviously shows that the system basically works. Loading data is more complex than extracting it, given that the **g_unit** values which tie the system together should never be manually input and unit names cannot be relied on to uniquely identify units. One working principle is that whenever a new unit is added to the system the same operation should create an ‘IsPartOf’ relationship to some higher level unit, so the unit can be identified by the combination of its own name and the higher unit’s name. This is most problematic for the combinations of parish names and ancient county names, and an additional **g_hint** value has been added to distinguish the small number of pairs of parishes within single counties that have the same name. These **g_hint** values will never be seen by users.
- Performance: The structure of the gazetteer makes even the simplest queries quite complex; for example, finding out which local government district a particular Civil Parish in a given ancient county was in requires seven distinct references to tables. However, the data model makes good use of the capabilities of relational database management systems and performance is better than expected. Some simple advance preparation of text strings has been done to improve performance: holding the assembled ‘authority’ for a given record with that record, rather than constructing it from one or more entries in the **g_authority** table plus other information on statutory orders, etc; and storing the preferred English name of each unit, status information and a label for the unit type in the **g_unit** table.
- Is it too flexible? Although the gazetteer tables are extensively constrained, the system is designed to accommodate various unusual relationships which are anything but the general rule (“Bishop’s Peculiars”, for example). A significant amount of checking code has already been written, including a requirement that only certain relationships may exist between specific combinations of types of unit (see appendix C). These additional checks cannot be implemented as constraints and are currently simply a set of commands is run each time the gazetteer is changed or extended, but

should in the longer term be implemented as triggers, small programs which run automatically each time certain changes are made to the gazetteer.

Data Entry Strategy

One reason it has taken us so long to get started on the construction of the gazetteer has been trying to work out how to handle the very complex information in Youngs' *Local Administrative Units*. The data model is capable of holding everything in Youngs, and the Isle of Wight system includes everything but the ecclesiastical hierarchy of units and parliamentary units, but the aim is to do as little work as possible directly on the parish entries:

- 1) Enter list of parishes from Youngs. The data transcribed will be limited to: name (and variant name); years of creation and abolition, if known; whether the parish was an Ancient Parish (or a township, chapelry or tything); whether the parish was a Civil Parish (and the date it gained that status); whether the parish was an Ecclesiastical Parish; plus notes and, for townships and chapelries, the name of the mother parish. This information is then used to create a number of records, including where relevant separate entries in *g_unit* for the Ancient/Civil Parish and the Ecclesiastical Parish, plus 'SucceededBy' relationships between the two and 'IsPartOf' relationships from both to the Ancient County, and from townships/chapelries to the mother parish.
- 2) Enter districts and hierarchic relationships from Youngs. Lists of the various kinds of districts in each county, and the parishes each one contained, appear in another part of Youngs which is easier to work with; the information on relationships between parishes and districts is effectively duplicated (and is sometimes inconsistent). We enter the lists of ancient districts (hundreds, boroughs), Poor Law Unions, Sanitary Districts and Local Government Districts, and the associated lists of the parishes within each one.
- 3) Improve coverage of registration units. Youngs does not cover Registration Districts, although these were largely identical with the Poor Law Unions which he does list, or the larger number of Registration sub-Districts. These were the main reporting units for demographic data between the 1840s and 1911, and therefore are key units within our statistical database. Broadly, we will use the 1911 census as our main authority for the names of these units.
- 4) Add boundary changes from census and vital registration reports. Much of the detailed information in Youngs' parish entries concern boundary changes, including those involved in the creation and abolition of units. Fortunately, the project already has very extensive transcriptions of similar but more detailed information from various official reports. The Isle of Wight gazetteer already includes such information from the 1891, 1901 and 1931 census reports. Using this information will both save a great deal of time and enrich the final product: these sources give precise calendar dates and a good deal of information about areas transferred, as well as details of the statutory orders making changes.

- 5) Adding additional information from Youngs. By this stage, without having done extensive transcriptions from the parish entries in Youngs, we will have replicated a large part of those entries. The final stage will involve a second pass through those entries, comparing our versions with his and adding additional information. Apart from information on the ecclesiastical and parliamentary hierarchies, this means, in practice, additional boundary changes which are almost entirely those affecting ecclesiastical parishes plus the footnotes giving Youngs' sources (mainly the *London Gazette*). Our Isle of Wight system includes these, but this final pass is likely to be time consuming, will certainly not be completed before the system goes live to the public, and may not be completed within our current funding.

The above is entirely concerned with Youngs, and therefore with England. Our sources for Wales and for Scotland will require somewhat different procedures. They generally provide simpler information than Youngs, and in particular say nothing about boundary changes. However, by taking these changes from official reports, we will be able to create a final resource which is substantially more uniform than the printed authorities that provide our core content. Replicating steps (1) and (2) for Scotland has already been done. Work on Wales would be greatly simplified by not including the very large number of townships listed by Melville Richards, which he gives no locations for and appear in none of our other sources.

The Descriptive Gazetteer

In addition to creating a systematic authority list/gazetteer for British administrative units, our commitments include computerising three gazetteers published in the late 19th century:

- John Bartholomew, *Gazetteer of the British Isles* (Edinburgh: Bartholomew, 1887) (918 pages)
- John Goring, *The Imperial Gazetteer of England and Wales* (Edinburgh: A Fullarton & Co., 6 vols., 1870-72) (2,391 pages)
- Frances H. Groome, *The Ordnance Gazetteer of Scotland* (Edinburgh: Thomas C. Jack, Grange Publishing Works, 6 vols., 1882-1885) (digitising this book is a collaboration with the 'Gazetteer for Scotland' project)

These books have many common features: each consists of alphabetically arranged entries; each entry begins with a **head-word**, i.e. the place-name usually in bold or upper case letters; it is followed by an indication of the **feature type** ('a parish', 'a river', etc) and then some indication of where the feature is, which almost always indicates a county, sometimes a relative location ('9 miles SW of Worcester') and never an absolute location such as latitude and longitude. The entry may or may not contain an extensive description of the place, or even a short history. If this material were being computerised in isolation, it could obviously be efficiently stored as a single table, searchable by head-word, possibly qualified by the county, and also through full text searching of all words in the entry.

Our system enables you to use the descriptive gazetteer in this way, but we have added considerable value by linking it with the administrative unit gazetteer. The two have to be kept distinct because they are concerned with fundamentally different kinds of entity, administrative units and ‘places’: a unit may contain a large number of ‘places’, while a place can give its name to several different units. One specific consequence of this is that the longer descriptive entries explicitly cover, in turn, a series of administrative units. For example, the *Imperial Gazetteer* entry for Ledbury, Herefordshire, describes it as ‘A town, a parish, a sub-district, and a district’ and consists of four distinct parts describing each of these entities in turn. In other words, a single entry covers three distinct polygons plus an undefined ‘town’.

As a result, descriptive gazetteer information is held in two separate tables, **g_text** and **g_text_gaz**, with links to the main gazetteer held in **g_text_link**. Each original entry forms the basis of a record in **g_text_gaz**, which includes the head-word, the feature-type and an ID number, but the actual description is held as one or more records in **g_text**. Simple searches of the descriptive gazetteer look at the **g_name** column in **g_text_gaz**, while full text searches use a special index on the **g_text** column of **g_text**.

As descriptive text is loaded into this structure, links are created with the main gazetteer:

All entries are linked to at least one county. Some entries describe the place as being ‘in’, or ‘on the borders of’ two counties, and both links are then created. Given that most English counties exist in three forms in the main gazetteer, as an Ancient, Registration and Administrative county, there is an obvious question whether all should be linked.

If the entry says it is a parish, or is in a particular parish, a link is created to that parish provided it can be found in the main gazetteer. In England, there is again the issue of whether to create links to the Ecclesiastical Parish or just the Ancient/Civil Parish. At present, *Imperial Gazetteer* entries are linked to both but Bartholomews only to the AP/CP.

Imperial Gazetteer entries often name the Registration District a unit was in. We also try to find a descriptive entry matching the manors in the main gazetteer, and are flexible on the kind of unit involved provided the name matches and is in the right parish.

Obviously, not all linked descriptive text is sensibly included in web pages but, for example, the links between descriptive entries and counties permits us to offer an alternative search mechanism in which the user specifies the county.

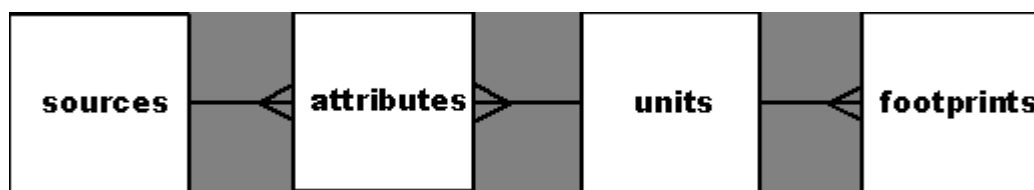
Note that although we are extracting the feature type information from the text, so that it appears in the web pages listing possible matching entries, it is not currently possible to specify the kind of object you are searching for. Enabling this will require us to impose a feature type thesaurus on the feature types, so that people looking for rivers can find entries for ‘streams’ and ‘affluents’, the latter term being commonly used in Bartholomew’s. This has yet to be attempted, but seems straightforward; the Alexandria Digital Library’s Feature Type Thesaurus looks appropriate.

Much more sophisticated searching would be possible if geographical coordinates could be associated with the entries, where these do not relate to administrative units we have already matched. We are considering matching our list against the Ordnance Survey 1:50,000 gazetteer, although this may have copyright implications.

In general, we have already taken the integration of the descriptive entries with the main gazetteer further than we expected.

Adding 'Footprints' and Statistics

The gazetteer is the core of the final system, but it needs to be linked to locational and attribute data. An earlier 'roadmap' for the project provided this overall schematic:



Simplified GBH: Entity-Relationship Diagram

The prototype gazetteer as a whole is the 'units' box here. 'Footprints' covers the existing GIS, currently held outside Oracle mainly in Arc/Info software. 'Attributes' mainly covers our statistical database but also the entries in the descriptive gazetteers and, in principle, the Travellers' Tales.

Footprints:

Adding these is conceptually straightforward: polygons will be extracted from our GIS software and added to the system using Oracle Spatial software, which enables polygons to be held as a column within an otherwise conventional table. This table will have a many-to-one relationship with the **g_unit** table, each administrative unit being associated with one or more polygons, allowing for different polygons covering different periods of time. Much work has been done over the last two years to create very large but otherwise conventional polygon coverages recording parish boundaries at each census date. Creating these from the original arc-based 'changing boundary' GIS, which remains problematic, has involved losing all the date stamp information, but by adding the raw boundary change data to the gazetteer we are effectively putting it back in.

We have already added all polygons we currently hold for the Isle of Wight. These include our own mapping of changing Civil Parishes, Local Government Districts and Poor Law Unions/Registration Districts (the island was a single Union/District for the entire period these units existed). We have also added polygons for Ancient Parishes as mapped by Roger Kain and Richard Oliver at Exeter University; unavoidably, these differ slightly from our own polygons for the same parishes, even where the actual boundaries were identical. We will be able to construct a number of other geographies, such as Ancient Counties and Hundreds, by assembling them from parishes and other units we hold.

Statistics:

There is no perfect solution for making available what is currently a very large database of historical statistics over the web. The problem is that the database currently consists of approaching 200 separate tables, and writing code to extract data from each of these separately would take too much time and, more importantly, create a system that was very hard to maintain. The many separate tables exists for the obvious reason that the data began, on paper, as many separate tables.

The prototype database does, however, incorporate a partial solution which will enable a substantial fraction of our statistical database to be made available via one single table. **g_data**, in which the actual numerical values are held in a single column. The other columns record:

- What geographical unit the data related to, via the **g_unit** ID values which are at the heart of the gazetteer.
- What date the data concern.
- What table within the census tables the data come from; this currently just says ‘Census of Population’ but another component of our overall system that is nearly complete is a list of all tables in all census reports.
- What the number measures. This is the problem area, as the meaning of many numbers in the more complex tables in the database is defined partly by the values in other columns. This is particularly true of occupational and cause of death tables, and tables based on complex cross-tabulations: age against sex against occupation against birthplace, for example. However, many values can be simply defined: total population, male engineering workers and so on. One of the changes we are making to the main database as we rebuild it on Tiger is a standard dictionary of column names, to be used in all tables, and this will identify which columns can be used in isolation, and so added to the web site.

Other data will be used in derived form, for example occupational data placed into a simpler set of categories; or included in separate thematic tables designed to drive mapping tools; or simply be downloadable as spreadsheets via the source documentation system.

To simplify linkage to the statistical data, a series of *views* have been defined which are in fact simply ways of seeing the current contents of the gazetteer but can be treated as tables containing lists of units of a given type at a given census date, as they would appear in the census report. For example, the view **g_lgd_1951** contains, for the Isle of Wight:

Local Govt. District	Type	Admin. County	ID
VENTNOR	UD	ISLE OF WIGHT	10000010
SANDOWN-SHANKLIN	UD	ISLE OF WIGHT	10000008
ISLE OF WIGHT	RD	ISLE OF WIGHT	10000005
COWES	UD	ISLE OF WIGHT	10000003

RYDE	MB	ISLE OF WIGHT	10000002
NEWPORT	MB	ISLE OF WIGHT	10000001

NB the ID values in the above example are not final. The code which creates this view is as follows:

```

create or replace view g_lgd_1951 as
select substr(n.g_name, 1, 30)    g_lgd,
       substr(s.g_status, 1, 4)   g_lgt,
       substr(rn.g_name, 1, 16)   g_admc,
       u.g_unit                   g_unit
from   g_unit u, g_name n, g_status s,
       g_rel r, g_unit ru, g_name rn
where  n.g_unit = u.g_unit and
       s.g_unit = u.g_unit and
       r.g_unit = u.g_unit and
       ru.g_unit = r.g_rel_to and
       rn.g_unit = ru.g_unit and
       u.g_unit_type = 'LG_DIST' and
       r.g_rel_type = 'IsPartOf' and
       ru.g_unit_type = 'ADM_CNTY' and
       (u.c_start_y <= 1951 or u.c_start_y is null) and
       (u.c_end_y > 1951 or u.c_end_y is null) and
       (s.c_start_y <= 1951 or s.c_start_y is null) and
       (s.c_end_y > 1951 or s.c_end_y is null) and
       (r.c_start_y <= 1951 or r.c_start_y is null) and
       (r.c_end_y > 1951 or r.c_end_y is null);

```

The final system

As already noted, the current web interface is designed to help the project team and various collaborators, especially those with local expertise, to check the contents. The final system is designed to serve a much wider audience:

- Search results returned in map form:** The current web interface returns the results of a search as a list of units, and tries to indicate which part of the country they were in by listing all the higher level units they were part of. This could maybe be simplified by listing only certain types of higher level unit, most obviously counties, but we hope to be able to return search results as maps of Great Britain with the different units shown by symbols; clicking on the symbol will select that unit. NB (a) this raises the issue of how to handle several units for the same place (see above); (b) a different kind of map would seem to be needed where the search was by location.
- A home page for each unit:** The current interface provides just one page for each unit. The final system will offer a number of pages per unit. Users will be taken initially to a ‘home page’ which provides access to other information, as discussed below, plus selected information from what will be presented in more detail on other pages. That summary *might* include: really basic information on the type of unit, and

its dates of creation and abolition, as at the top of the current page; a small map showing location as a point on an outline map of Britain; a slightly larger map showing the unit's boundaries superimposed on a scanned historical map of the area; some basic statistics, such as a population time series.

- **Access to historic maps:** We can offer larger excerpts from Ordnance Survey one-inch maps, designed to fill the user's screen, from both the mid-19th century *First Series* and the late 1940s *New Popular* edition.
- **Access to statistical information:** As discussed above, the gazetteer will be directly linked to a substantial subset of our statistical database which will include almost all data held for parishes and much of our district level data. More of an issue is how to present this information graphically.
- **A glossary of terms:** A trivial enhancement to the current system, which is already provided for via the *notes* columns in the tables defining types of units and possible status values, is including explanatory text for each category of unit. The SCAN project has agreed that we may make use of the glossary they have already written describing different types of Scottish unit.
- **A gateway to other web sites:** One aim of the system is to help users who do not know much about an area's detailed administrative history to find the names of administrative units they can look up in other reference resources. Where those other resources are database-driven web sites, direct links should be possible. Our demonstration system written two years ago included a link to the National Register of Archives' ARCHON system, and this at very least should be included in the final system.

Appendix A: Types of Units

As defined in the table **g_unit_type**, the following values of *g_unit_type* can exist within the gazetteer:

Type	ID	Level	Meaning
PAR_UNIT	0	7	Parish-level Unit
ROOT	1	0	World
STATE	2	1	State
NATION	3	2	Nation
ADM_CNTY	4	4	Administrative County
ANC_CNTY	5	4	Ancient County
PR_CNTY	6	4	Poor Law/Registration County
ANC_DIST	7	5	Ancient District
PR_DIST	8	5	Poor Law Union/Reg. District
SAN_DIST	9	6	Sanitary District
SUB_DIST	10	6	Registration sub-District
LG_DIST	11	6	Local Government District
MANOR	12	8	Manor
SCO_CNTY	31	4	Scottish County
SCO_PAR	33	7	Scottish Parish
SCO_BURGH	34	8	Scottish Burgh
IRL_PROV	41	3	Irish Province
IRL_CNTY	42	4	Irish County

The ID value above provides the most significant part of the *g_unit* ID values which identify the various units within the gazetteer, so 4791 will be a parish (ID=0), 9000056 will be a Sanitary District and 1000001 identifies the world as a whole (NB there is no intention to extend the system beyond the British Isles, but that term is probably best avoided).

As defined in the table **g_unit_type_level**, each possible value of *g_unit_type* is assigned to one of the following geographical levels, from 0 meaning the whole world to 8 meaning sub-parish units:

0	World	5	District 1
1	State	6	District 2
2	Nation	7	Parish
3	Region	8	sub-Parish
4	County		

Appendix B: Possible Status Values

As defined in the table **g_status_type**, the possible values of *g_status* are as follows. NB to avoid ambiguity, each value can be used only with units of a specific type:

Unit Type	Status	Meaning
ANC_DIST:	Borough	Borough
	Hundred	Hundred
	Liberty	Liberty
LG_DIST:	CB	County Borough (i.e. City of London)
	CC	County Corporate
	LB	London Borough
	MB	Municipal Borough
	RD	Rural District
	UD	Urban District
PAR_UNIT:	AP	Ancient Parish
	Ch	Chapelry
	CP	Civil Parish
	EP	Ecclesiastical Parish
	ExP	Extra-Parochial Area
	Tg	Tything
	Tn	Township
PR_DIST:	Inc	Incorporation
	PLPar	Poor Law Parish
	PLU	Poor Law Union
	<i>RegD</i>	<i>Registration District</i>
SAN_DIST:	RSD	Rural Sanitary District
	USD	Urban Sanitary District
SCO_BURGH:	<i>LBu</i>	<i>Large Burgh</i>
	<i>PBu</i>	<i>Police Burgh</i>
	<i>RBu</i>	<i>Royal Burgh</i>
	<i>SBu</i>	<i>Small Burgh</i>

Italicised terms have yet to be used.

Appendix C: Possible Relationships between Units

As defined in the table **g_legal_rel**, the following values of *g_rel_type* are allowed between units of the specified types:

Unit	Relationship	Related Unit	Unit	Relationship	Related Unit
ADM_CNTY	AbolishedToCreate	ADM_CNTY	PAR_UNIT	IsPartOf	ANC_DIST
ADM_CNTY	AbolishedToEnlarge	ADM_CNTY	PAR_UNIT	IsPartOf	LG_DIST
ADM_CNTY	BoundaryChange		PAR_UNIT	IsPartOf	PAR_UNIT
ADM_CNTY	Event		PAR_UNIT	IsPartOf	PR_DIST
ADM_CNTY	IsPartOf	NATION	PAR_UNIT	IsPartOf	SAN_DIST
ADM_CNTY	ReducedToCreate	ADM_CNTY	PAR_UNIT	IsPartOf	SUB_DIST
ADM_CNTY	ReducedToEnlarge	ADM_CNTY	PAR_UNIT	ReducedToCreate	PAR_UNIT
ANC_CNTY	IsPartOf	NATION	PAR_UNIT	ReducedToEnlarge	PAR_UNIT
ANC_CNTY	SucceededBy	ADM_CNTY	PAR_UNIT	SucceededBy	PAR_UNIT
ANC_CNTY	SucceededBy	PR_CNTY	PR_CNTY	SucceededBy	ADM_CNTY
ANC_DIST	IsPartOf	ANC_CNTY	PR_DIST	AbolishedToCreate	PR_DIST
ANC_DIST	SucceededBy	LG_DIST	PR_DIST	AbolishedToEnlarge	PR_DIST
PR_CNTY	AbolishedToCreate	LG_CNTY	PR_DIST	BoundaryChange	
PR_CNTY	AbolishedToEnlarge	LG_CNTY	PR_DIST	Event	
PR_CNTY	BoundaryChange		PR_DIST	IsPartOf	PR_CNTY
PR_CNTY	Event		PR_DIST	ReducedToCreate	PR_DIST
PR_CNTY	IsPartOf	NATION	PR_DIST	ReducedToEnlarge	PR_DIST
PR_CNTY	ReducedToCreate	LG_CNTY	SAN_DIST	IsPartOf	PR_CNTY
PR_CNTY	ReducedToEnlarge	LG_CNTY	STATE	IsPartOf	ROOT
LG_DIST	AbolishedToCreate	LG_DIST	SUB_DIST	AbolishedToCreate	SUB_DIST
LG_DIST	AbolishedToEnlarge	LG_DIST	SUB_DIST	AbolishedToEnlarge	SUB_DIST
LG_DIST	BoundaryChange		SUB_DIST	BoundaryChange	
LG_DIST	Event		SUB_DIST	Event	
LG_DIST	IsPartOf	ADM_CNTY	SUB_DIST	IsPartOf	PR_DIST
LG_DIST	ReducedToCreate	LG_DIST	SUB_DIST	ReducedToCreate	SUB_DIST
LG_DIST	ReducedToEnlarge	LG_DIST	SUB_DIST	ReducedToEnlarge	SUB_DIST
MANOR	IsPartOf	PAR_UNIT	SCO_CNTY	IsPartOf	NATION
NATION	IsPartOf	STATE	SCO_PAR	IsPartOf	SCO_CNTY
PAR_UNIT	AbolishedToCreate	PAR_UNIT	IRL_PROV	IsPartOf	NATION
PAR_UNIT	AbolishedToEnlarge	PAR_UNIT	IRL_CNTY	IsPartOf	STATE
PAR_UNIT	BoundaryChange		IRL_CNTY	IsPartOf	NATION
PAR_UNIT	Event		IRL_CNTY	IsPartOf	IRL_PROV
PAR_UNIT	IsPartOf	ANC_CNTY			

Appendix D: Unit Name Statuses and Languages

Each unit name within the gazetteer must be assigned to one of the following status categories, as defined in the **g_name_status** table:

Status	Meaning	Notes
P	Preferred	One and only one required per unit, except where names exist in more than one language in which case one and only one preferred name per language is required.
O	Official	A name which was at some date the official name of the unit, but is not now the preferred name. Only one such name may exist in any one language at any one date.
A	Alternate	
B	Abbreviation	These are needed, mainly for counties, to enable the gazetteer to link to certain census sources.

This system is a little complex so as to meet the expectations of both librarians, expecting a single preferred name, and archivists expecting to know the changing official names over time. The **g_name** table does hold dates, but these only appear when official names other than the preferred name appear.

Every unit name within the gazetteer must be assigned to one of the following languages, as defined in the **g_name_language** table:

Status	Meaning	Notes
C	Cornish	
E	English	
F	French	
G	Gaelic	i.e. Scottish Gaelic.
I	Irish	i.e. Erse
M	Manx	
S	Scotch	
W	Welsh	

NB at present only English and Welsh are actually used. Obviously, to make the gazetteer truly multi-lingual alternative web interfaces in each language need to be created.

Appendix E: Column names within the Gazetteer

A standard set of column names is used in all tables within the gazetteer, as follows:

Column name	Type	Meaning	Appears in:
<i>c_end</i>	Date	End date.	<i>g_name, g_rel, g_status, g_unit</i>
<i>c_end_d</i>	Number	Day part of end date.	<i>g_name, g_rel, g_status, g_unit</i>
<i>c_end_m</i>	Number	Month part of end date.	<i>g_name, g_rel, g_status, g_unit</i>
<i>c_end_y</i>	Number	Year part of end date.	<i>g_name, g_rel, g_status, g_unit</i>
<i>c_start</i>	Date	Start date.	<i>g_name, g_rel, g_status, g_unit</i>
<i>c_start_d</i>	Number	Day part of start date.	<i>g_name, g_rel, g_status, g_unit</i>
<i>c_start_m</i>	Number	Month part of start date.	<i>g_name, g_rel, g_status, g_unit</i>
<i>c_start_y</i>	Number	Year part of start date.	<i>g_name, g_rel, g_status, g_unit</i>
<i>c_year</i>	Number	Year	<i>g_data, g_foot</i>
<i>d_cen1</i>	Number	Date of first census for which statistics for the area transferred is given.	<i>g_rel</i>
<i>d_cen2</i>	Number	Date of second census for which statistics for the area transferred is given.	<i>g_rel</i>
<i>d_pop1</i>	Number	Population of the area transferred at the date of the first census.	<i>g_rel</i>
<i>d_pop2</i>	Number	Population of the area transferred at the date of the second census.	<i>g_rel</i>
<i>demo_flag</i>	Number	Has value 1 if the record should appear in the demonstration web site.	<i>g_foot, g_name, g_rel, g_status, g_text_link, g_unit</i>
<i>g_auth</i>	Text	Code identifying the authority for the information.	<i>g_text_gaz</i>
<i>g_auth_note</i>	Text (max. 400ch.)	Additional details of the authority. NB depending on the authority this may contain specific information, such as a page number.	<i>g_auth</i>
<i>g_authority</i>	Text (max. 20ch.)	Code identifying the authority for the information.	<i>g_auth, g_authority, g_data, g_name, g_order, g_rel, g_status, g_unit</i>
<i>g_data</i>	Number	Statistical data value, such as 'Total Population'.	<i>g_data</i>
<i>g_east</i>	Number	Easting part of grid reference. Derived from a unit's footprint, or the footprint of associated units, and may not be that precise.	<i>g_unit</i>
<i>g_foot</i>	Spatial object	Polygon representing the unit's boundary at a given date.	<i>g_foot</i>
<i>g_hint</i>	Number	Generally null, but is used to distinguish otherwise identical units, most obviously parishes with the same name in the same county.	<i>g_unit</i>

Column name	Type	Meaning	Appears in:
<i>g_link_status</i>	Text (max. 2ch.)	One-letter code indicating the status of a link between the descriptive gazetteer and an admin unit: M=Main, S=Subsidiary	g_text_link
<i>g_name</i>	Text (max. 40ch.)	Name by which an administrative unit was known. The version in g_unit is the preferred name in English, held there to improve performance.	g_name g_text_gaz g_unit
<i>g_name_language</i>	Text (max. 2ch.)	The language of the name.	g_name, g_name_language
<i>g_name_status</i>	Text (max. 2ch.)	One-letter code defining the status of a unit name, I.e. preferred, alternate, etc.	g_name, g_name_status
<i>g_north</i>	Number	Northing part of grid reference. Derived from a unit's footprint, or the footprint of associated units, and may not be that precise.	g_unit
<i>g_order</i>	Text (max. 400ch.)	Number of a statutory order implementing a boundary change, held as text.	g_order
<i>g_part</i>	Text (max. 80ch.)	Details of area transferred through boundary changes.	g_rel
<i>g_pt_area</i>	Number	Area of an area transferred by a boundary change, in acres.	g_rel
<i>g_rel_to</i>	Number	Number uniquely identifying the secondary administrative unit involved in a relationship. This will be null where the record covers an event affecting just the primary unit.	g_rel
<i>g_rel_type</i>	Text (max. 20ch.)	Type of relationship/event.	g_legal_rel, g_rel, g_rel_type
<i>g_rel_unit_type</i>	Text (max. 10ch.)	Type of the second unit in a relationship.	g_legal_rel
<i>g_seq</i>	Number	Number uniquely identifying the row within a table, to which an authority entry links.	g_auth, g_name, g_rel, g_status, g_text, g_text_link
<i>g_status</i>	Text (max. 10ch.)	The status of the administrative unit. The version held in g_unit holds all the different values for a unit held in g_status as a single string, separated by '/'.	g_status g_status_type g_unit
<i>g_text</i>	CLOB	The actual text from a descriptive gazetteer.	g_text
<i>g_text_id</i>	Number	Number uniquely identifying an entry in the descriptive gazetteer; primary key for g_text_gaz.	g_text, g_text_gaz, g_text_link
<i>g_text_key</i>	Number	Number uniquely identifying a separate part of a descriptive entry; primary key for g_text.	g_text
<i>g_type_label</i>	Text (max. 40ch.)	Label giving full name for a g_unit_type value.	g_unit, g_unit_type
<i>g_type_level</i>	Number	Number identifying the geographical level of a particular type of unit.	g_unit, g_unit_type, g_unit_type_level
<i>g_type_num</i>	Number	Number uniquely identifying a unit type. These numbers form part of the core 'g_unit' ID values.	g_unit_type
<i>g_unit</i>	Number	Number uniquely identifying each administrative unit. The six least significant digits are a sequence number for a given type of unit; the remaining digits identify the type of unit.	g_auth, g_data, g_foot, g_name, g_rel, g_status, g_text_link, g_unit

Column name	Type	Meaning	Appears in:
<i>g_unit_type</i>	Text (max. 10ch.)	Text string identifying a type of unit.	g_legal_rel, g_status_type, g_text, g_text_gaz, g_unit, g_unit_type
<i>g_unit_type_text</i>	Text (max. 40ch.)	The actual text from a descriptive gazetteer entry describing the type of entry.	g_text_gaz
<i>g_xref</i>	Number	A value of g_text_id, which links one entry in the descriptive gazetteer to another based on 'See ...' references in the original.	g_text_gaz
<i>h_hous1</i>	Number	Number of inhabited houses within the area transferred, at the date of the first census.	g_rel
<i>h_hous2</i>	Number	Number of inhabited houses within the area transferred, at the date of the second census.	g_rel
<i>meaning</i>	Text (max. 200ch.)	Explanation of a code number or abbreviation.	g_authority, g_name_language, g_name_status, g_order, g_rel_type, g_status_type, g_unit_type_level
<i>notes</i>	Text (max. 2000ch.)	Comments.	g_auth, g_legal_rel, g_name, g_rel, g_rel_type, g_status, g_status_type, g_unit, g_unit_type
<i>rev_meaning</i>	Text (max. 40ch.)	Plain text statement of the converse relationship within the g_rel table, i.e. the relationship that the unit related to has with the primary unit. For example, the relationship 'AbolishedToCreate' becomes 'was created through the abolition of'.	g_rel_type
<i>sort_order</i>	Number	Number ensuring records appear in correct order in output.	g_name_status
<i>table_name</i>	Text (max. 40ch.)	The name of a table within the gazetteer.	g_auth
<i>var_name</i>	Text (max. 20ch.)	Name of variable. NB likely to be replaced by more complex description of data.	g_data