Creating DDI Metadata for a Diverse Body of Aggregate Statistics:
Experience with the GB Historical GIS

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

Great Britain Historical GIS Project:
A Vision of Britain though Time

University of Portsmouth

Great Britain Historical GIS

- Project running since 1994
- Integrating GB historical statistics from 1801
  - Spatial framework recording changing systems of counties, districts and parishes
  - Supporting diverse research
    - HALCyon – Healthy Aging across the Life Course
    - Work for Dept of Environment, Farming & Rural Affairs
  - Popular web site: www.visionofbritain.org.uk
    - c. 80,000 unique users per month
  - Data downloads via UKDA, EDINA
- Presentation focuses on work since 2004

GBH GIS Content: Census Statistics

- Primary focus on recurring tables with detail below county level:
  - Almost every parish-level table: 1801, 1831, 1851-1961
  - District-level age structure 1851-1971 (5-year bands)
  - Occupations: 1841, 1881, 1921, 1931, 1951
  - Social Class: (1841-1931), 1951, 1961
  - Housing: 1901, 1931, 1951, 1961, 1971
  - Education: 1851, 1951, 1961
  - Birthplaces: not yet in site
  - Boundary changes
- Coverage of Scotland not as systematic
- Almost no current coverage of Ireland

Great Britain Historical GIS Project:
A Vision of Britain though Time

University of Portsmouth

GBH GIS Content: Vital Statistics

- Total Births, Marriages and Deaths plus IM since 1840s
- Annual data by district 1911-1974 for c. 1,500 districts: births, deaths, infant mortality, perinatal mortality, stillbirths
- Decennial cause of death data 1851-1910
  - Data mostly from Bob Woods (Liverpool)
  - Gender by age by cause of death for c. 630 districts
  - Five different cause of death classifications over six decades
  - 25-30 causes in each classification
  - Mapping to standard classification developed by Graham Mooney (Johns Hopkins)

GBH GIS Content: Other Statistics

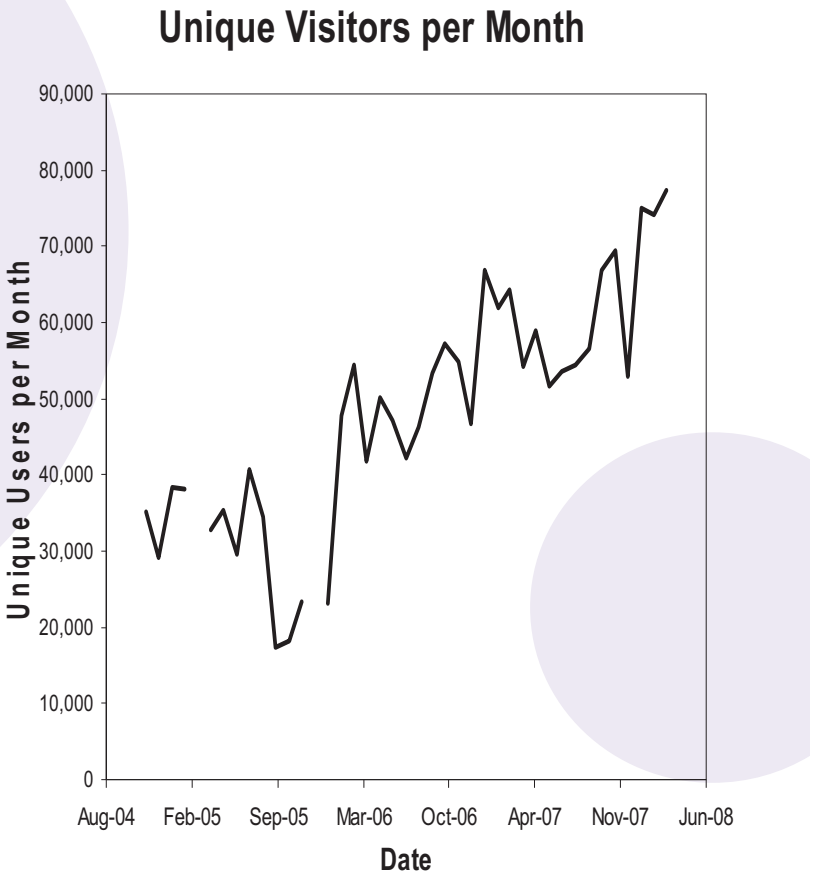
- Economic Distress
  - Poor Law, trade union and other data
  - Earliest part of collection, now being added to main system
    - Ministry of Labour *Local Unemployment Index 1927-39*
    - 1945-72 unemployment statistics from *Labour Gazette*
    - Wage Statistics by occupation and town 1851-1904
- Agriculture and Land Use
  - 1801 Crop Returns, by parish
  - County Farm Census for at least every 10th year 1869-1970
  - Land Utilisation Survey of Great Britain (1930s)
- Electoral Statistics
  - Votes for each party in each constituency in each election 1833-2008

What's different about the GBH GIS

- System NOT required to support data download
  - UKDA and EDINA already did this
- System IS required to support on-line visualisation as maps and especially as local time series
  - i.e. Needed to be able to extract data from original context in particular surveys
  - Longer-term goal of enabling very large scale analysis
- System must allow new data on new topics to be added without adding database tables
  - Necessary because our capacity to write code limited
- So all statistics held in one column of one table
  - So this table currently has 11,736,994 rows

Vision of Britain: Monthly Usage

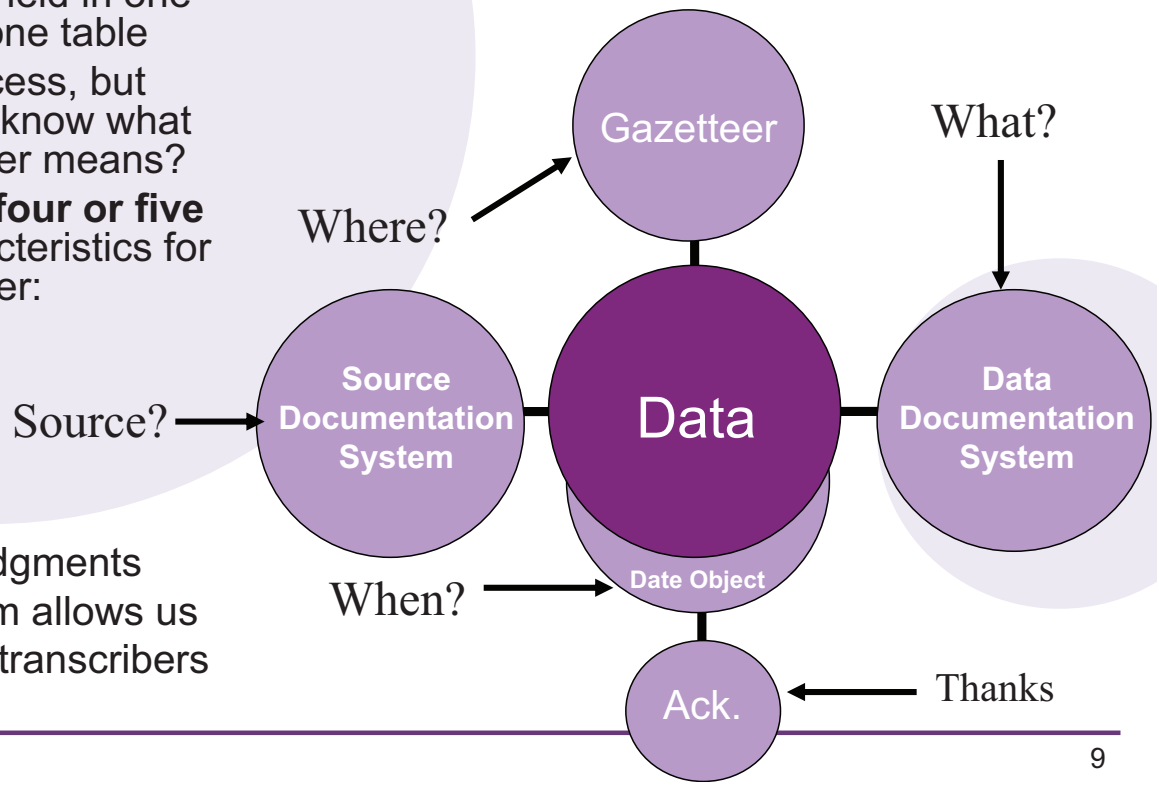
- Site launched October 2004
  - Many problems
- Site greatly extended October 2005
- Re-launched spring 2009
  - Sorry the graph has not been updated!
- Most users arrive from a search engine and land on a place-specific page



What the papers say about Vision of Britain:

GBH GIS Architecture Overview

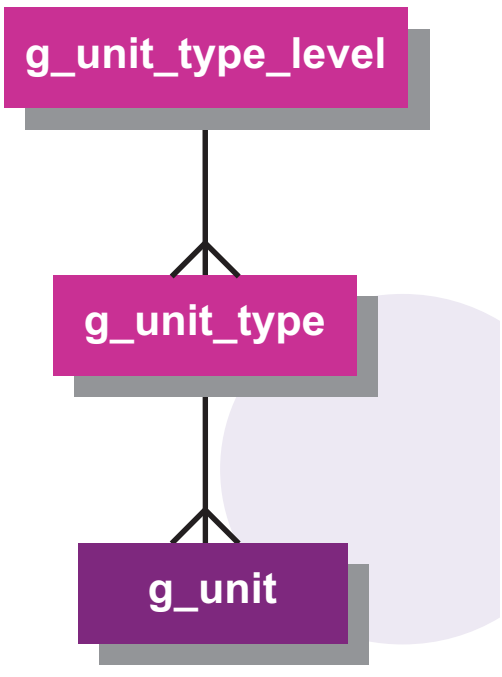
- Data-centric, not polygon-centric
- All 11,736,994 data values are held in one column of one table
- Easy to access, but how do we know what each number means?
- We record **four or five** main characteristics for each number:



Acknowledgments sub-system allows us to identify transcribers

AU Ontology: In the beginning was the unit …

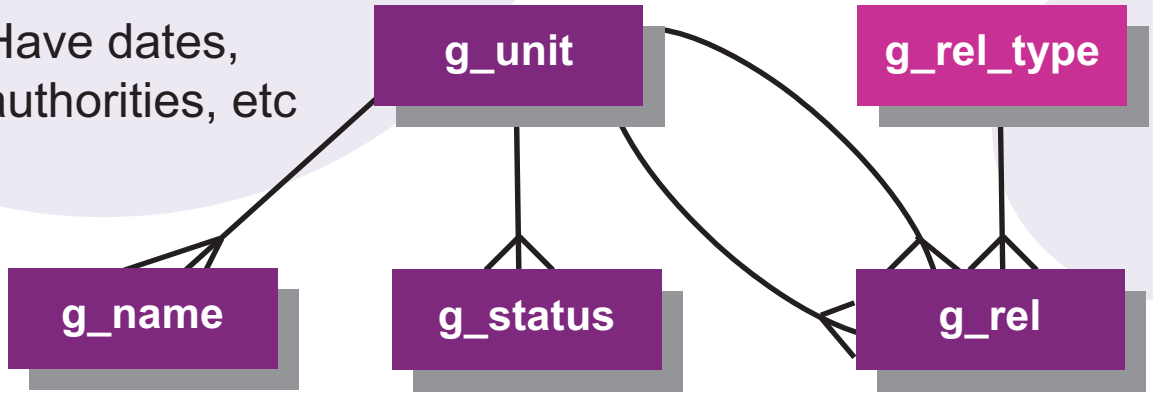
- Administrative areas, so corporate bodies with legally-defined boundaries, and dates of creation and abolition
  - Districts and Unitary Authorities
  - Hundreds and Wapentakes
  - States of Europe since 1815
- Basic unit record is minimal: ID number, type, dates of existence, and immediate and ultimate authorities
- All units are assigned to a type, such as *Ancient County* or *Sanitary District*, and types are assigned to one of 13 geographical levels, e.g. *County*



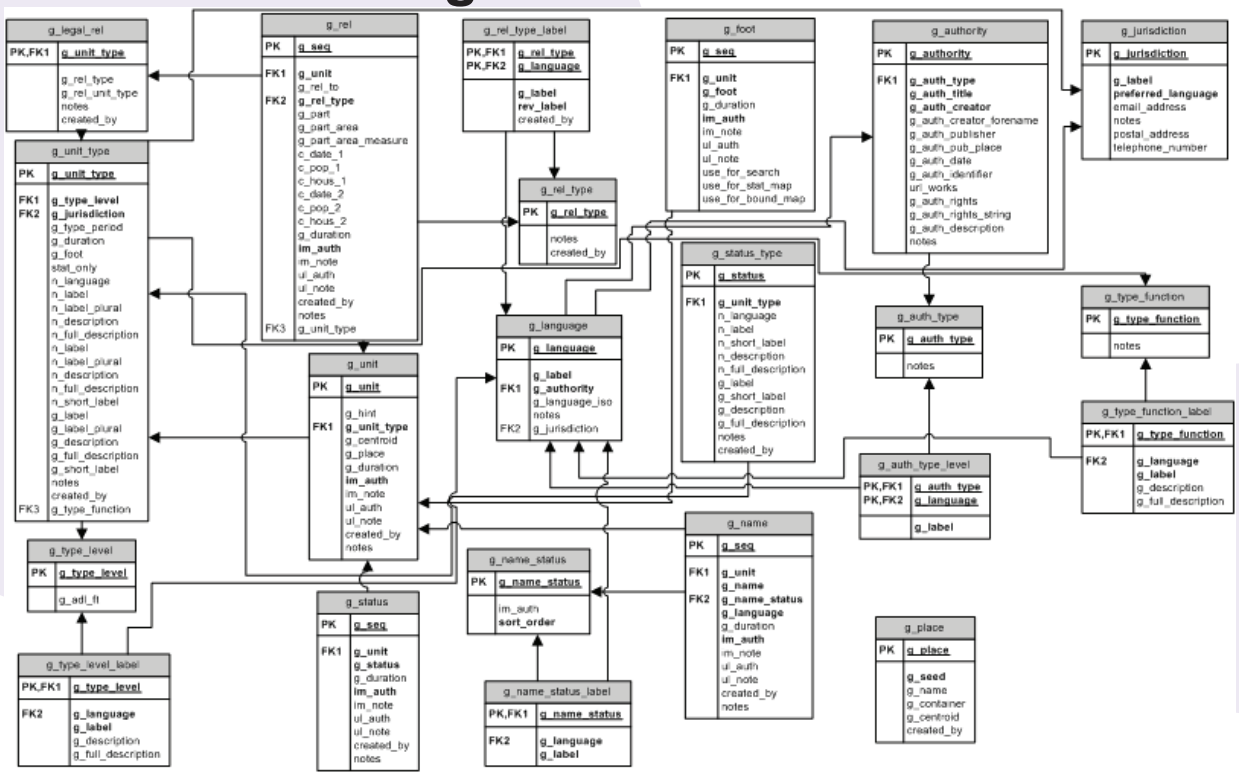
Unit relationships

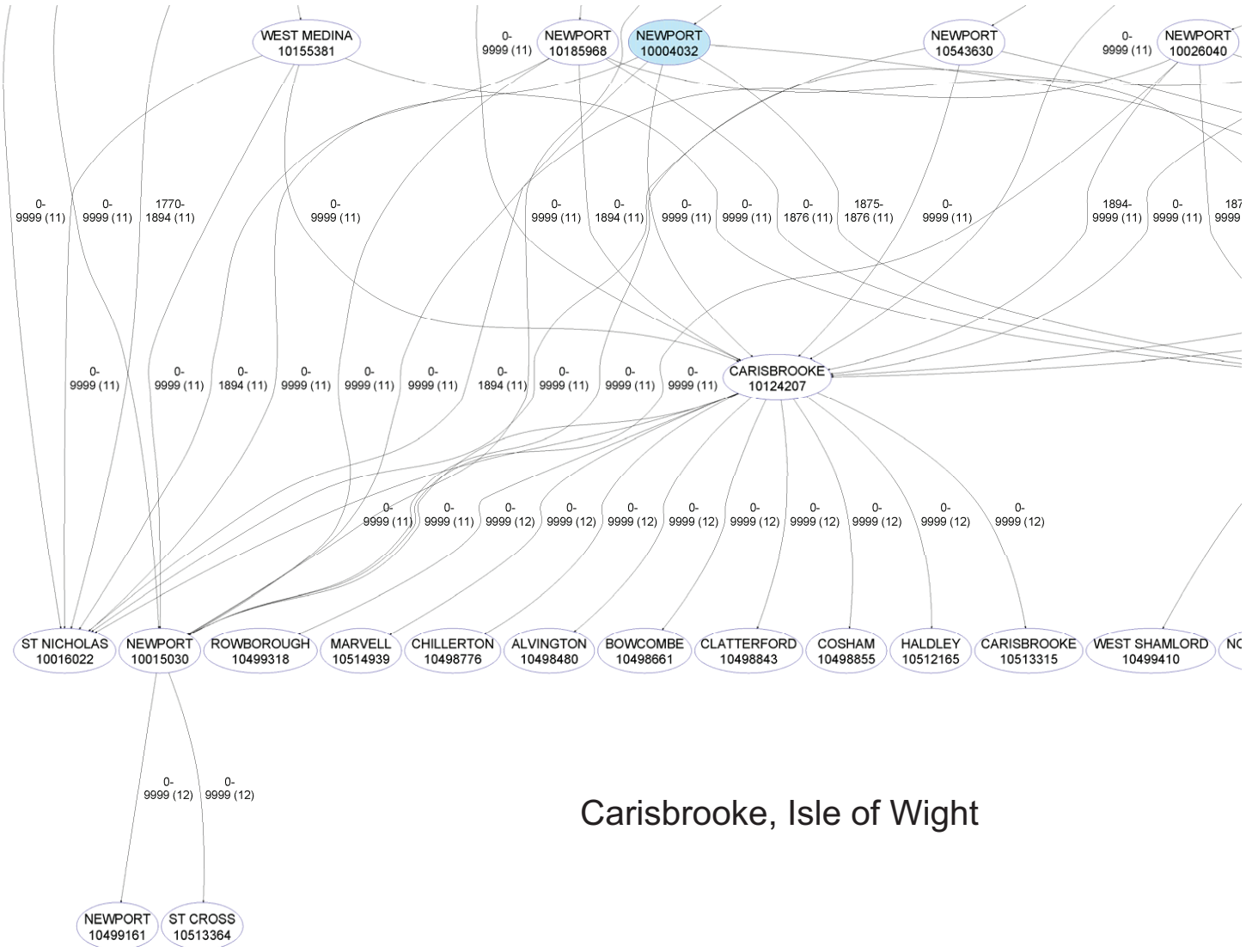
- All held in single table, allowing many-to-many relationships
- Current system has 78,471 units but 249,241 relationships
- Have dates, authorities, etc

- IsPartOf
- SucceededBy ('see also')
- AdministeredBy
- Boundary Changes
  - ReducedToEnlarge
  - ReducedToCreate
  - AbolishedToEnlarge
  - AbolishedToCreate
  - BoundaryChange (other unit unknown)



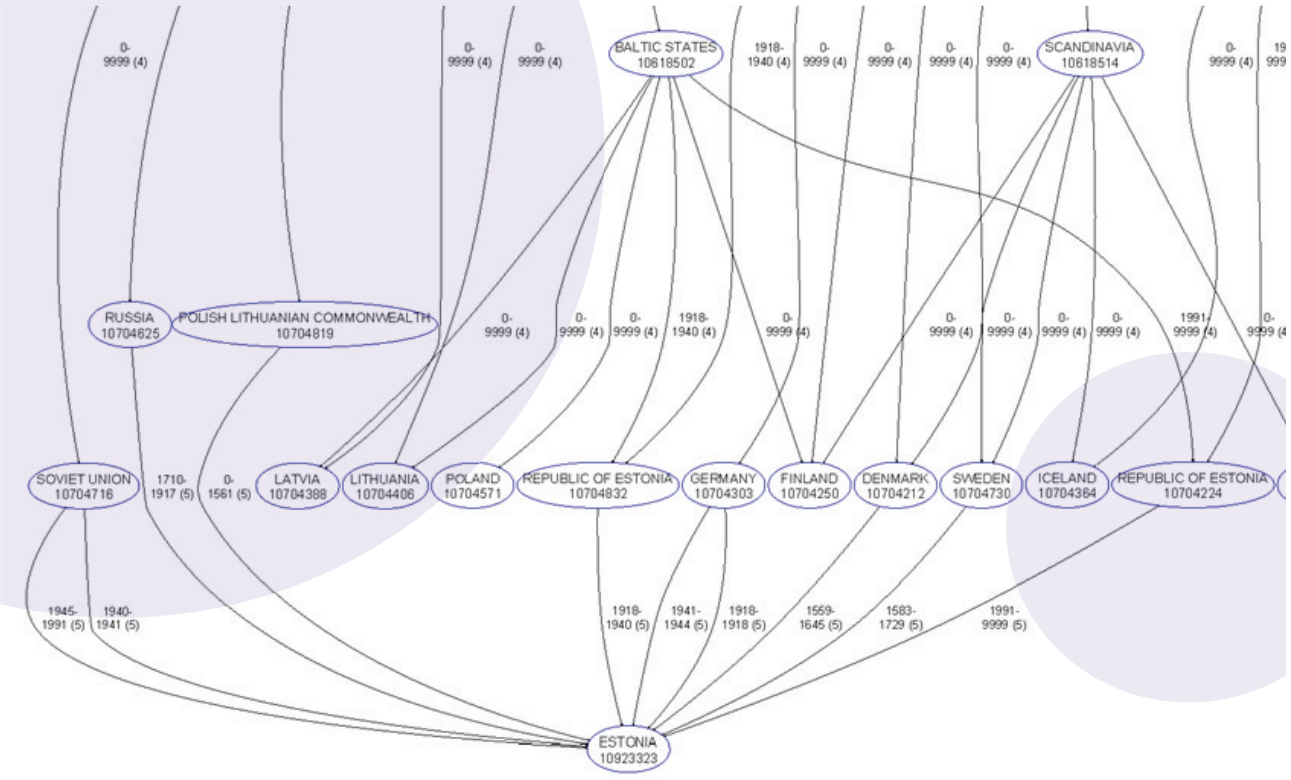
EHGIS – E-R Diagram





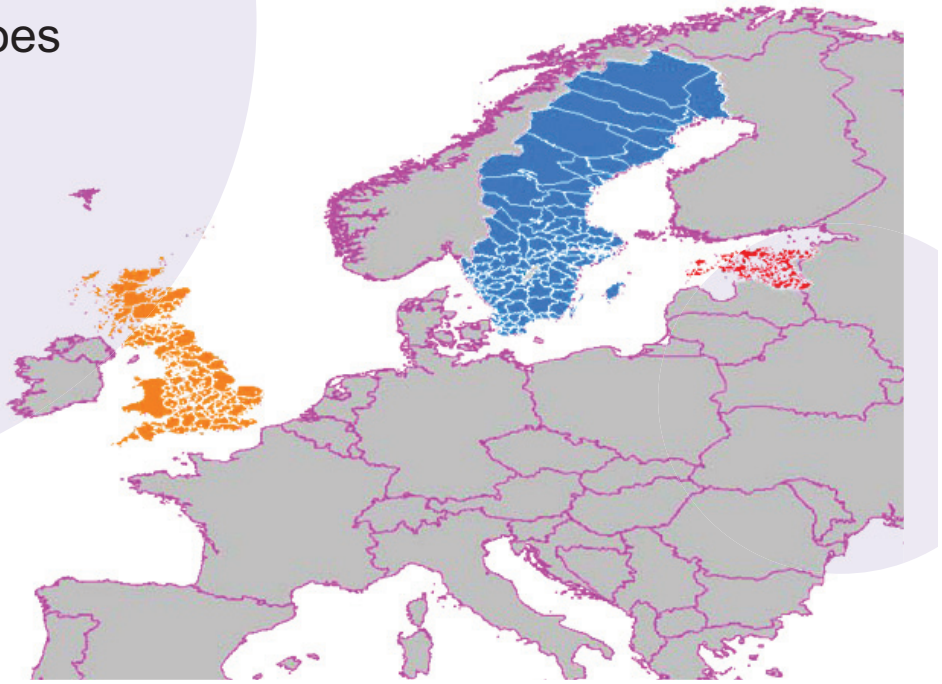
Carisbrooke, Isle of Wight

Estonia



Boundary Mapping: Britain, Estonia, Sweden

- Example shows boundaries down only to county-level
- Current system goes down to or below parishes for all 3 countries
- 75,291 units and 75,986 polygons
  - Many units have no polygons
- Largest multi-national historical GIS?



Recording When: Date Objects

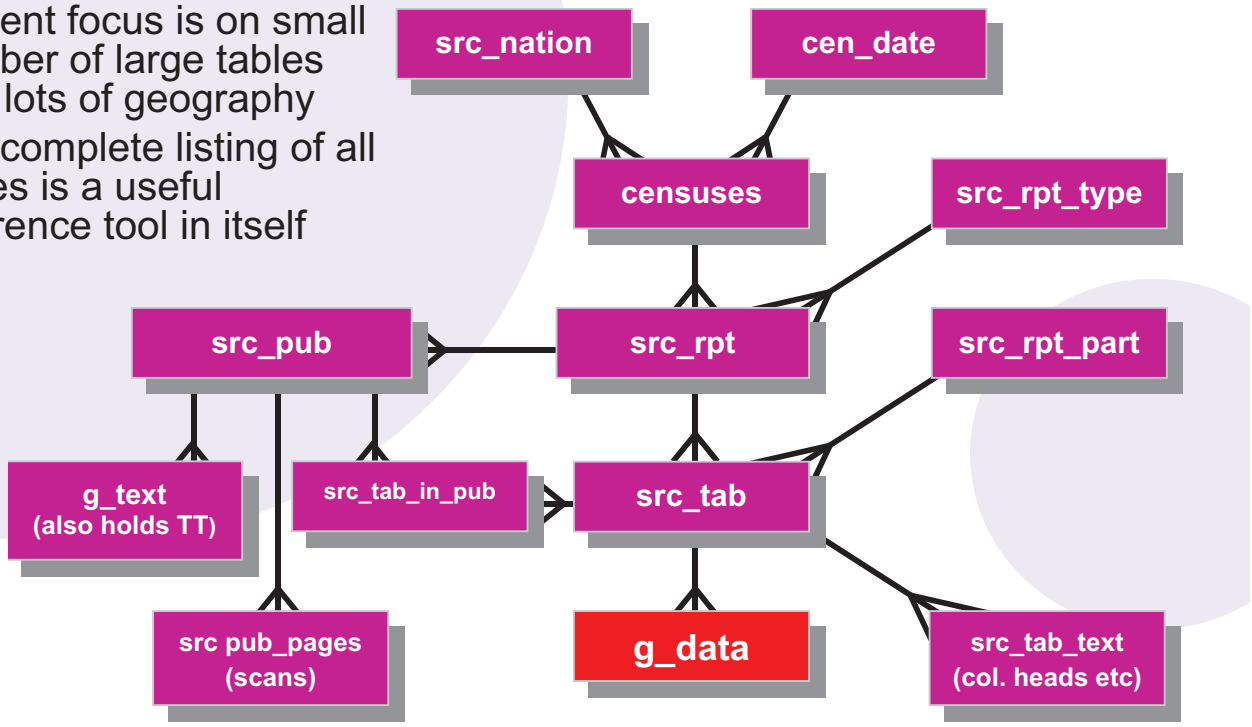
- This is the *fairly* easy one
- Dates held directly within the data table
- … but we store dates in a specially defined *date object* (using Oracle's object-relational capabilities), not in a conventional Oracle date value:
  - Census data stored using simple year value
    - Means graphs, etc, treat censuses as regularly spaced
    - Exact date of each census stored elsewhere in system
  - Unemployment counts, etc, stored using full calendar dates
  - Vital registration data stored usually with a duration, I.e. a pair of full calendar dates
  - Elsewhere, we store text strings within date objects, such as "around the reign of Edward II", but with statistical data we require at least a specific year

Recording Source: The Source Documentation System

- Records:
  - All the censuses taken in Britain between 1801 and 1961
  - All the reports published by each census
  - All the tables published in each report
    - The SDS creates a unique ID for each census table
  - For tables computerised by the project, column and row headings, and notes applying to the table as a whole
- This information enables us to create web pages reconstructing the original tables, correctly labelled
  - System provides 'drill-down' mechanism, rather than exact reconstruction
  - Follows hierarchies of units recorded in the gazetteer
  - Reconstructions can include data values not covered by the DDS
- Like the gazetteer, a reference resource in its own right
  - System holds text of the Preliminary and some General Reports
  - Could also hold a complete set of image scans of the reports
  - Records many-to-many relationships between tables and pages
- Non-census sources identified by simpler Dublin Core-based system

- Could hold and reconstruct GB census reports *in toto*
- Current focus is on small number of large tables with lots of geography
- Our complete listing of all tables is a useful reference tool in itself

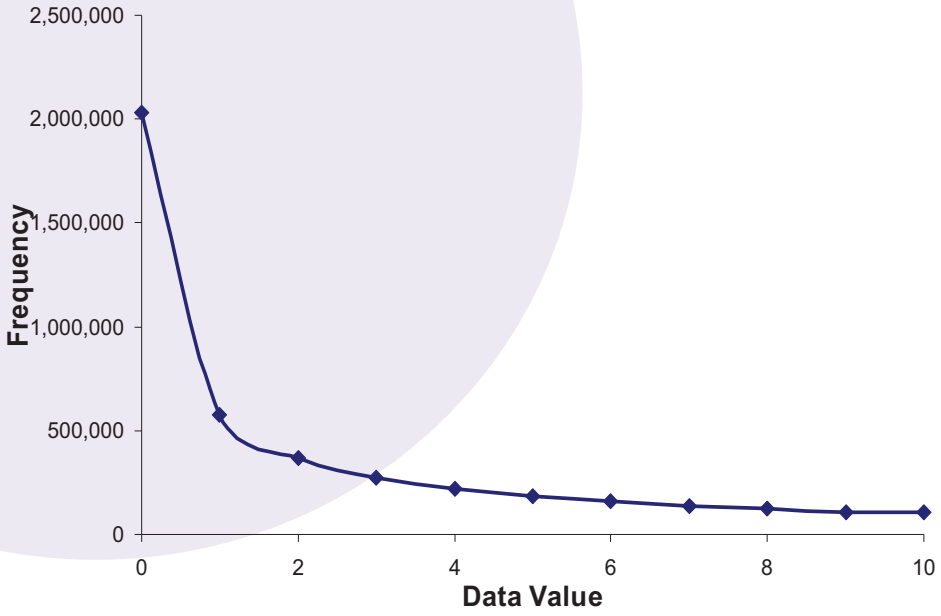
SDS Schematic:



Recording What: The Data Documentation System

- A relational implementation of the Data Documentation Initiative's Aggregate/Tabular Data extension (DDI 2, extended)
  - XML-based standard supported by most major data archives, tho' this extension not yet widely implemented
- Defines meaning of each data value by location within an *nCube* whose dimensions consist of *variables*, each defined as a set of *categories*
  - For example, gender by age by cause of death
  - Each different categorisation of cause of death is a *different* variable
    - The six Registrar General's *Decennial Supplements* 1851-1901 use five different cause of death classifications, which we place in a *variable group*
  - *Category groups* enable mapping of one variable into another
- Each nCube also belongs to a *universe*
  - Two nCubes might both combine gender and age, but one covers total population, the other total deaths
- Supports very general approach to creating maps and graphs
  - Choropleth maps and graphs of pre-computed rates are most accessible
  - Rules for graphing nCubes, e.g. 2D: n x 2 categories = 'population pyramid'
  - Mapping nCubes permits v large number of logically valid maps

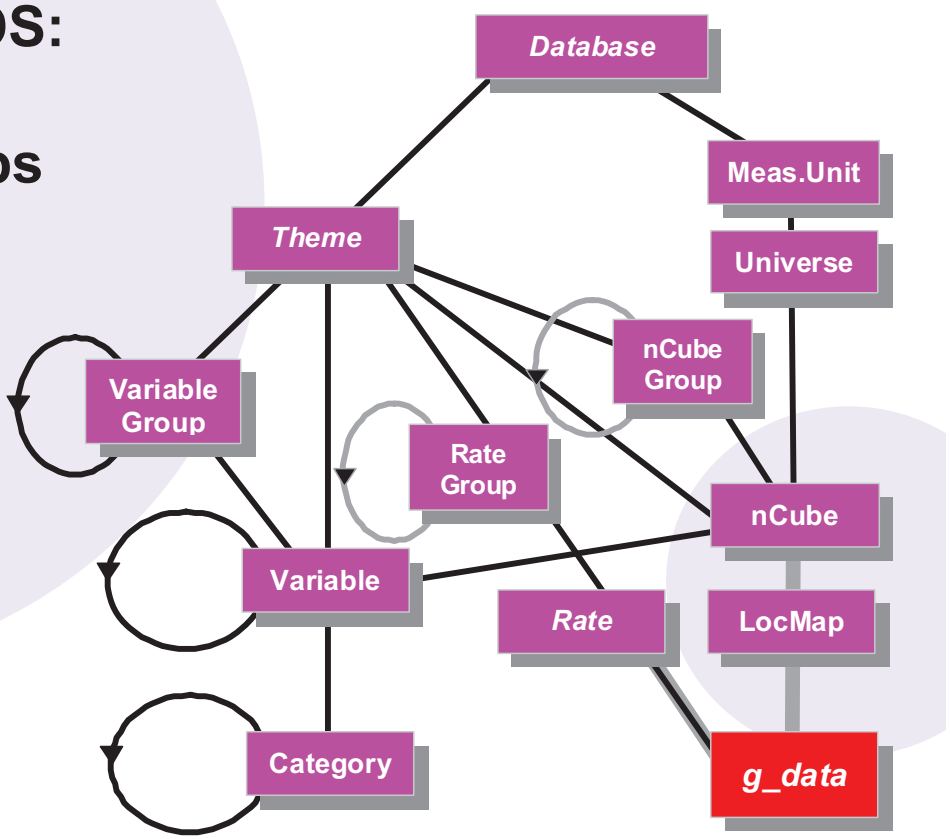
Frequency of Data Values 0-10 in GBH GIS Data Table



- Yes, we store 2,029,002 copies of the number "0"
- We record the meaning of statistics **entirely** by metadata

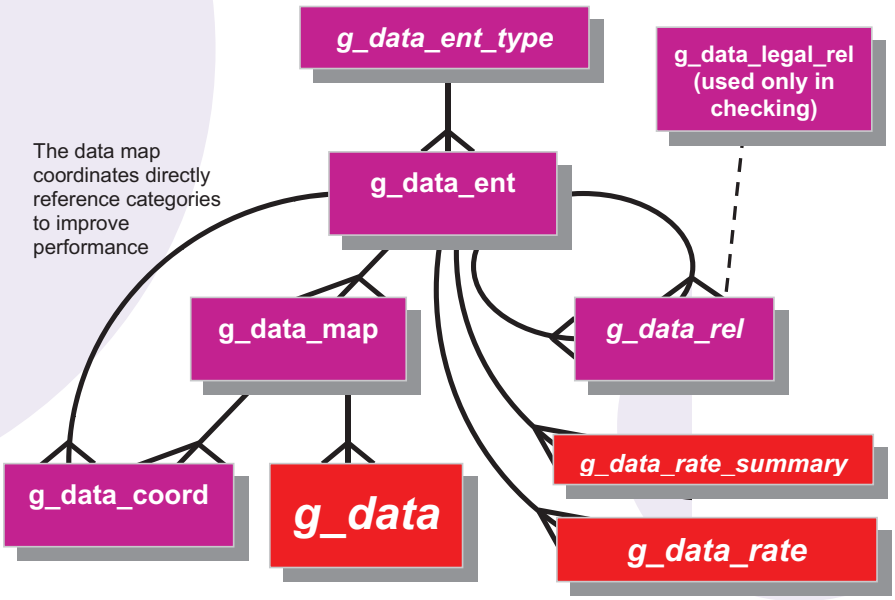
GBH GIS DDS: Entities and Relationships

- but this is not the database structure ...



- Data model further abstracts the DDI
- Holding all labels and text for all entities in one table simplifies searching
- System extensively denormalised for performance
  - e.g. universes and measurement units are held both as entities *and* as nCube attributes
  - data_rate and its summary hold all data needed for most accessible maps and graphs
- Fixed set of tables holds any amount of data in any number of nCubes, with any number of dimensions

GBH GIS DDS Schematic:

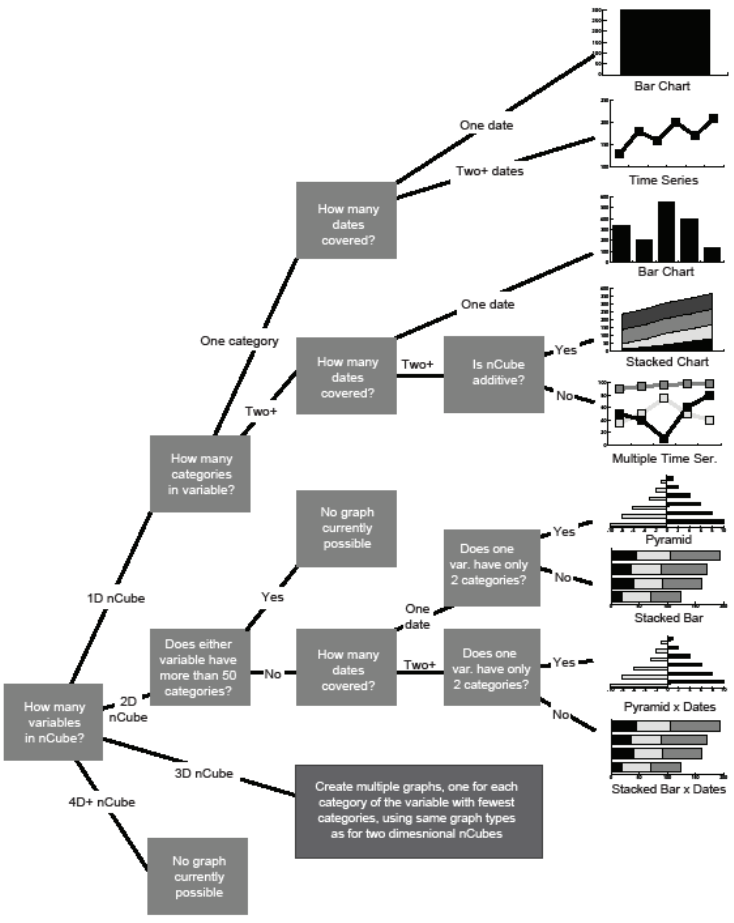


Driving analysis

- DDS defines new data that *can* be generated
  - New simplified nCubes defined by variable-to-variable mappings
  - Rates, constructed from available data values within nCubes
  - To come: redistrictings, based on holding geography conversion table data as relationships within gazetteer
- The system *does* pre-generate all this
  - As statistics are loaded into system, all derivable data values are derived, to accelerate web site
  - Example: decennial mortality:
    - Graham Mooney mapped all causes of death in the five classifications 1851-1910 to single simplified classification
    - Mappings defined within DDS as category-to-category links
    - Single query then generated c. 800,000 new derived values
    - "Bloody hell that was quick!" (G. Mooney, pers. com., 2005)
  - Currently derived data obvious, but massive and widely disseminated

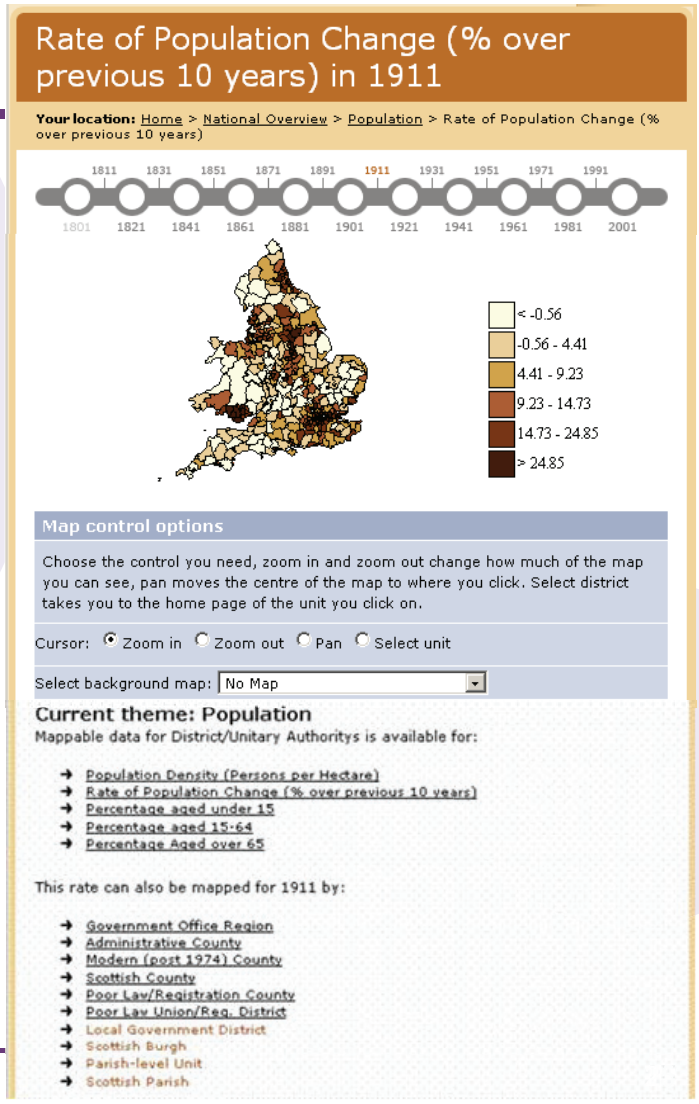
Decision tree for generating graphs from nCubes

- Graphs generated very conventional
- Higher-order nCubes cannot be graphed
- … but neither limitation is fundamental to system
- Please fund a more advanced visualisation project!



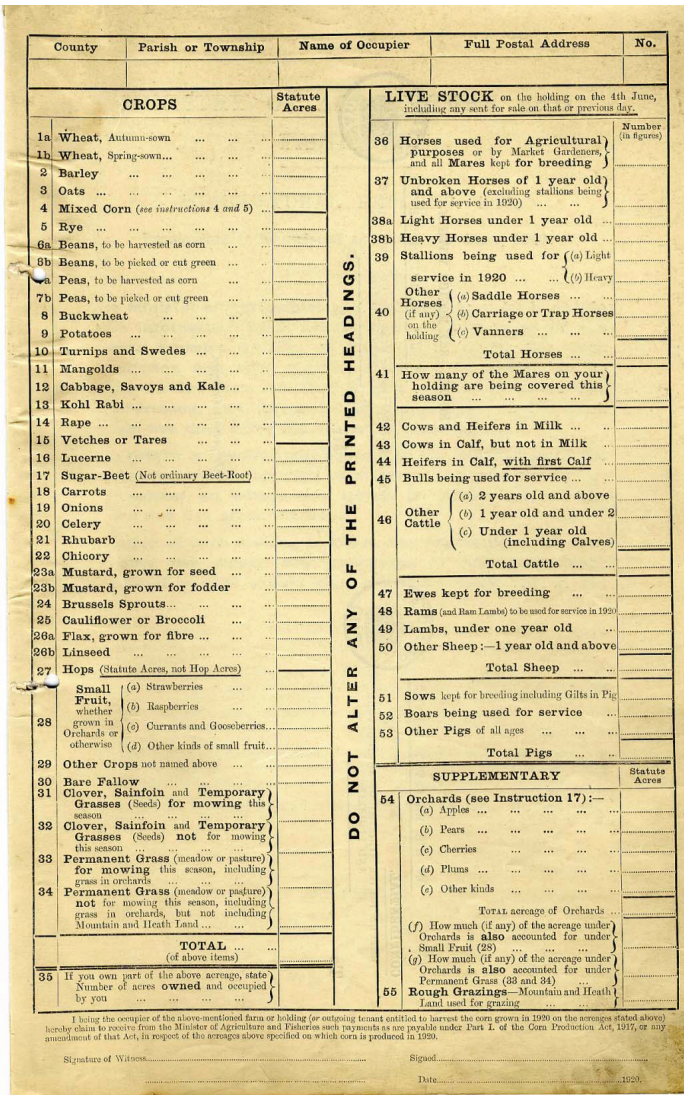
Generating maps from Vision of Britain

- Here again, a very conventional graphic
- But note the range of options:
  - Zoom in and out
  - Change date (via timeline bar)
  - Change variable
  - Change type of unit
  - Change backcloth



Agricultural Census Form for 1922

- NB Pre-computer farm level microdata preserved for 1941 only


County | Parish or Township | Name of Occupier | Full Postal Address | No.

CROPS — Statute Acres

1a Wheat, Autumn-sown
1b Wheat, Spring-sown
2 Barley
3 Oats
4 Mixed Corn (see instructions 4 and 5)
5 Rye
6a Beans, to be harvested as corn
6b Beans, to be picked or cut green
7a Peas, to be harvested as corn
7b Peas, to be picked or cut green
8 Buckwheat
9 Potatoes
10 Turnips and Swedes
11 Mangolds
12 Cabbage, Savoys and Kale
13 Kohl Rabi
14 Rape
15 Vetches or Tares
16 Lucerne
17 Sugar-Beet (Not ordinary Beet-Root)
18 Carrots
19 Onions
20 Celery
21 Rhubarb
22 Chicory
23a Mustard, grown for seed
23b Mustard, grown for fodder
24 Brussels Sprouts
25 Cauliflower or Broccoli
26a Flax, grown for fibre
26b Linseed
27 Hops (Statute Acres, not Hop Acres)
28 Small Fruit, whether grown in Orchards or otherwise: (a) Strawberries (b) Raspberries (c) Currants and Gooseberries (d) Other kinds of small fruit
29 Other Crops not named above
30 Bare Fallow
31 Clover, Sainfoin and Temporary Grasses (Seeds) for mowing this season
32 Clover, Sainfoin and Temporary Grasses (Seeds) not for mowing this season
33 Permanent Grass (meadow or pasture) for mowing this season, including grass in orchards
34 Permanent Grass (meadow or pasture) not for mowing this season, including grass in orchards, but not including Mountain and Heath Land
TOTAL (of above items)
35 If you own part of the above acreage, state Number of acres owned and occupied by you

DO NOT ALTER ANY OF THE PRINTED HEADINGS.

LIVE STOCK on the holding on the 4th June, including any sent for sale on that or previous day. — Number (in figures)

36 Horses used for Agricultural purposes or by Market Gardeners, and all Mares kept for breeding
37 Unbroken Horses of 1 year old and above (excluding stallions being used for service in 1920)
38a Light Horses under 1 year old
38b Heavy Horses under 1 year old
39 Stallions being used for service in 1920: (a) Light (b) Heavy
40 Other Horses (if any) on the holding: (a) Saddle Horses (b) Carriage or Trap Horses (c) Vanners
Total Horses
41 How many of the Mares on your holding are being covered this season
42 Cows and Heifers in Milk
43 Cows in Calf, but not in Milk
44 Heifers in Calf, with first Calf
45 Bulls being used for service
46 Other Cattle: (a) 2 years old and above (b) 1 year old and under 2 (c) Under 1 year old (including Calves)
Total Cattle
47 Ewes kept for breeding
48 Rams (and Ram Lambs) to be used for service in 1920
49 Lambs, under one year old
50 Other Sheep:—1 year old and above
Total Sheep
51 Sows kept for breeding including Gilts in Pig
52 Boars being used for service
53 Other Pigs of all ages
Total Pigs

SUPPLEMENTARY — Statute Acres

54 Orchards (see Instruction 17):—
(a) Apples
(b) Pears
(c) Cherries
(d) Plums
(e) Other kinds
Total acreage of Orchards
(f) How much (if any) of the acreage under Orchards is also accounted for under Small Fruit (28)
(g) How much (if any) of the acreage under Orchards is also accounted for under Permanent Grass (32 and 34)
55 Rough Grazings—Mountain and Heath Land used for grazing

I being the occupier of the above-mentioned farm or holding (or outgoing tenant entitled to harvest the corn grown in 1920 on the acreages stated above) hereby claim to receive from the Minister of Agriculture and Fisheries such payments as are payable under Part I. of the Corn Production Act, 1917, or any amendment of that Act, in respect of the acreages above specified on which corn is produced in 1920.

Signature of Witness ... Signed ...
... Date ... 1920.

A vision of BRITAIN THROUGH TIME

A vision of Britain between 1801 and 2001. Including maps, statistical trends and historical descriptions.

Help using this website

Home | Places | Statistical atlas | Historical maps | Census reports | Travel writing | Learning resources

Home / Statistical Atlas

Statistical Atlas

The Vision of Britain system contains historical statistics from two centuries, plus the boundaries of the main statistical reporting unit. This national overview gives you direct access to a vast statistical atlas of Britain, organised by theme.

Population
Densities and rates of growth. Age structure: concentrations of young people, and of the old.

Industry
How people made their living: identifying farming communities, pit villages, mill towns and so on.

Social structure
Social class and socio-economic groups. Concentrations of professionals, and of unskilled labourers.

Language & learning
How many people were at school. The distribution of graduates, and of the unqualified.

Agriculture & Land Use
How we use the land; crops and farm animals

Births & deaths
Numbers of babies born. Deaths, especially infant deaths. Causes of death by age and gender.

Employment & poverty
Unemployment rates and poor relief. What proportions of men and of women did paid work?

Housing
Numbers of houses. Overcrowding: persons per room. Amenities: central heating, baths and WCs.

Roots & religion
Numbers going to church, and the kinds of churches they went to. The different religions of modern Britain.

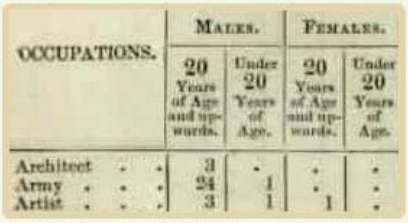
Politics
Election results, plus information on turn-out.

Home | Places | Statistical atlas | Historical maps | Census reports | Travel writing | Learning resources

Home / Expert search / Data Documentation

The Vision of Britain system holds a large library of historical statistics. We store all the data values in one big list, and for each value we record:

- WHERE the data value is for, by links to our administrative gazetteer
- WHEN the value is for, by a year or period
- The SOURCE, linking to our listings of Census Reports

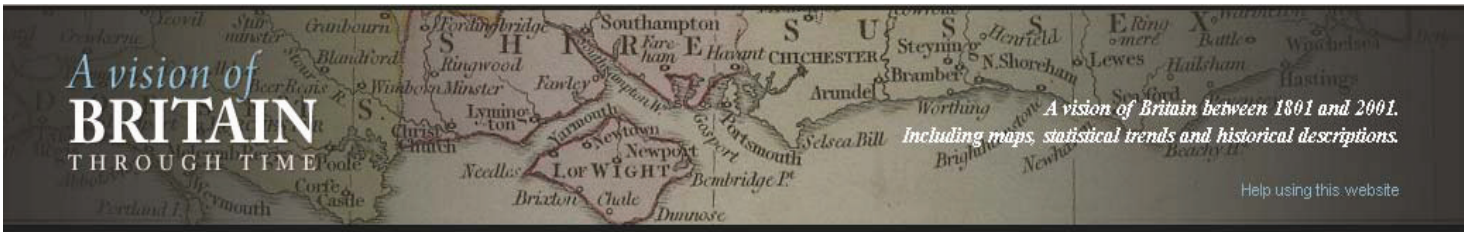


OCCUPATIONS.	MALES.		FEMALES.	
	20 Years of Age and upwards.	Under 20 Years of Age.	20 Years of Age and upwards.	Under 20 Years of Age.
Architect	3	.	.	.
Army	24	1	.	.
Artist	3	1	1	.

We record WHAT each number measures via our implementation of the Aggregate/Tabular Data Extension developed by the Data Documentation Initiative. This places each data value within an nCube whose dimensions are defined by variables (sometimes organised into variable groups), like occupation and gender and age, each of which is a set of categories, such as "Males" and "Females":

The implementation is structured as an ontology containing 6,152 entities of the following types, linked by 19,427 relationships:

Entity type	Entities in system
Databases	1
Themes	10
Rates	63
nCube Groups	12
nCubes	184
Variable Groups	16
Variables	169
Categories	5,623
Measurement Units	14
Universes	60



Home | Places | Statistical atlas | Historical maps | Census reports | Travel writing | Learning resources

Home / Expert search / Data Documentation / Themes

Themes are used in the Vision of Britain system to organise all our statistical data into a small number of broad categories, often identified by icons on the web site.

The following Themes are currently defined:

Entity ID	Entity Name
T_HOUS	Housing
T_IND	Industry
T_LAND	Agriculture and Land Use
T_LEARN	Learning and Language
T_POL	Political Life
T_POP	Population
T_REL	Roots and Religion
T_SOC	Social Structure
T_VITAL	Life and Death
T_WK	Work and Poverty

Data definition

Name:	Agriculture and Land Use
Identifier:	T_LAND
Type:	Theme (T)
Text:	This aspect of the Vision of Britain system presents data from two kinds of source: Agricultural Censuses, which gathered data on the kinds of crops planted and animals raised, usually from farmers; and Land Use Mapping, which actually surveyed land use, field by field. Work on this theme has been funded by DEFRA, the Environment Agency and the Frederick Soddy Trust. Our most detailed information on land use is, in fact, the complete set of the maps published by the Land Utilisation Survey of Great Britain, included in our 'Historical Mapping' section.

Theme "T_LAND" is contained within:

Databases:

Entity ID	Entity Name
DB	VoB Database

Theme "T_LAND" contains:

Rates :

Entity ID	Entity Name
R_LAND_BUILD	Land occupied by Buildings
R_LUSGB_unprod	Land Agriculturally Unproductive
R_LUSGB_arable	Percentage Arable (Land Utilisation Survey)
R_AGCEN_arable	Percentage Arable (Farm Census)
R_ORCHARD	Orchards as a percentage of total area
R_WHEAT	Wheat Farming
R_POTATO	Potato Farming

Entity ID	Entity Name
N_AGCEN_AREA_A	Total areas of reporting units
N_LAND2001	Land Use Statistics for 2001
N_LAND2001_TOT	Total area, in hectares
N_LAND2001_SIMP	Simplified Land Use Statistics for 2001
N_LUSGB	Land Utilisation in the 1930s (LUSGB data)
N_LUSGB_TOT	Total area, in acres
N_LUSGB_FARM	Farmland Use, as reported by the Land Utilisation Survey
N_AGCEN_LAND	Basic land use categories, as reported by the Agricultural Census
N_AGCEN_FARM	Farmland Use, as reported by the Agricultural Census
N_CROP_GEN	Generalised Crop Acreages
N_CROP1801	1801 Crop Acreages
N_CROP1801_TOTAL	1801 Crop Return Total Acreage
N_AGCEN_VEG_S_1961	Acreages of different vegetable, as reported by the 1961 Scottish Farm Census
N_AGCEN_SMFRUIT_S_1961	Acreages of different small fruit, as reported by the 1961 Scottish Farm Census

Variable Groups :

Entity ID	Entity Name
VG_LAND	Land Use
VG_GRASS	Grassland
VG_AGCEN_CROP	Crops
VG_ANIMALS	Animals
VG_FARMSIZE	Farm Size
VG_AGCEN_WORKERS	Farmworkers

Variables :

Entity ID	Entity Name
V_AGCEN_TENURE	Agricultural Census Tenure-type Categories for 1891
V_AGCEN_VEG_S_1961	Scottish Farm Census Vegetable Categories for 1961
V_AGCEN_SMFRUIT_S_1961	Scottish Farm Census Small Fruit Categories for 1961

Data definition

Name:	Animals
Identifier:	VG_ANIMALS
Type:	Variable Group (VG)
Text:	This variable group brings together the various classifications of types of animal, as used over the years by the Agricultural Census. Note that variable groups covering the different classifications of horse, of cows and so on are nested within this.

Variable Group "VG_ANIMALS" is contained within:

Themes:

Entity ID	Entity Name
T_LAND	Agriculture and Land Use

Variable Group "VG_ANIMALS" contains:

Variable Groups :

Entity ID	Entity Name
VG_AGCEN_HORSES	Horses
VG_AGCEN_CATTLE	Cattle
VG_AGCEN_SHEEP	Sheep
VG_AGCEN_PIGS	Pigs
VG_POULTRY	Poultry

Variables :

Entity ID	Entity Name
V_AGCEN_GOATS_1961	Agricultural Census Goat Categories for 1961

Data definition

Name:	Sheep
Identifier:	VG_AGCEN_SHEEP
Type:	Variable Group (VG)
Text:	This variable group brings together the different classifications of sheep used over the years by the Agricultural Census.

Variable Group "VG_AGCEN_SHEEP" is contained within:

Variable Groups:

Entity ID	Entity Name
VG_ANIMALS	Animals

Variable Group "VG_AGCEN_SHEEP" contains:

Variables :

Entity ID	Entity Name
V_AGCEN_SHEEP_GEN	Agricultural Census Generalised Sheep Categories
V_AGCEN_SHEEP_1866	Agricultural Census Sheep Categories for 1866-1891
V_AGCEN_SHEEP_1900	Agricultural Census Sheep Categories for 1900-10
V_AGCEN_SHEEP_1920	Agricultural Census Sheep Categories for 1920-30
V_AGCEN_SHEEP_1940	Agricultural Census Sheep Categories for 1940-50
V_AGCEN_SHEEP_1960	Agricultural Census Sheep Categories for 1960-70
V_AGCEN_SHEEP_S_1961	Scottish Farm Census Sheep Categories for 1961

Variable "V_AGCEN_SHEEP_1960" is contained within:

nCubes:

Entity ID	Entity Name
N_AGCEN_SHEEP_1960	Types of sheep as reported by the Agricultural Census.

Variable Groups:

Entity ID	Entity Name
VG_AGCEN_SHEEP	Sheep

Variables:

Entity ID	Entity Name
V_AGCEN_SHEEP_GEN	Agricultural Census Generalised Sheep Categories

Variable "V_AGCEN_SHEEP_1960" contains:

Categories :

Entity ID	Entity Name
C_AGCEN_SHEEP_1960_1	One year and over: Ewes kept for breeding
C_AGCEN_SHEEP_1960_2	One year and over: Two tooth (shearling) ewes to be put to the ram
C_AGCEN_SHEEP_1960_3	One year and over: Rams kept for service
C_AGCEN_SHEEP_1960_4	One year and over: Draft and cast ewes
C_AGCEN_SHEEP_1960_5	One year and over: Wethers and others
C_AGCEN_SHEEP_1960_6	Under one year

Data definition

Name:	One year and over: Wethers and others
Identifier:	C_AGCEN_SHEEP_1960_5
Type:	Category (C)
Text:	One year and over: Wethers and others. 'a castrated male sheep' (Source: English Nature, The upland management handbook, 'Glossary' (2001). A glossary on the BBC web site defines wethers more precisely as 'male sheep castrated at an early age before secondary sexual characters have developed'.

Category "C_AGCEN_SHEEP_1960_5" is contained within:

Variables:

Entity ID	Entity Name
V_AGCEN_SHEEP_1960	Agricultural Census Sheep Categories for 1960-70

Categories:

Entity ID	Entity Name
C_AGCEN_SHEEP_GEN_3	All Other Sheep

Category "C_AGCEN_SHEEP_1960_5" contains no lower-level entities.

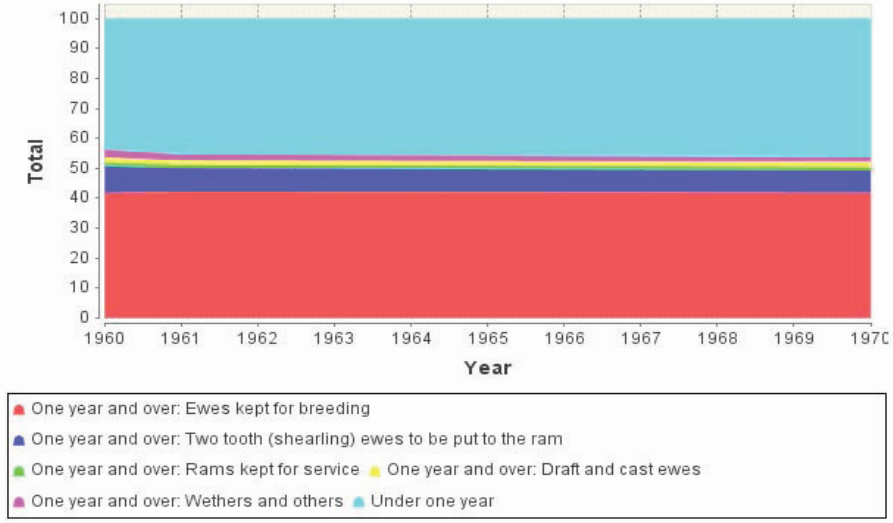
Administrative unit
Devon
Administrative County

Historical statistics
Population
Industry
Social Class
Learning and Language
Agriculture & Land Use
Life and Death
Work and Poverty
Housing
Boundary Map
Unit history & boundary changes
Census reports

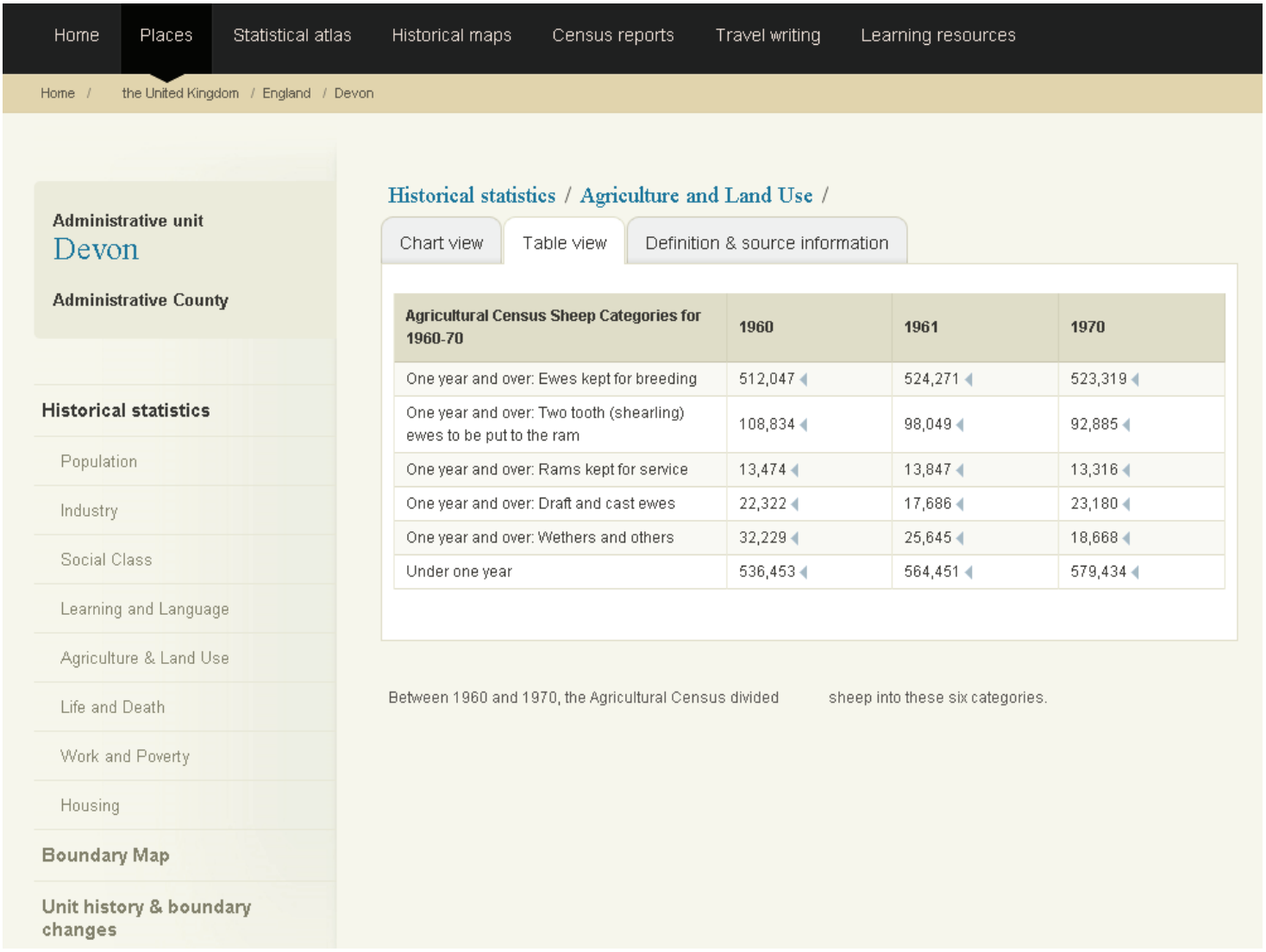
Historical statistics / Agriculture and Land Use /

Chart view | Table view | Definition & source information

Types of sheep as reported by the Agricultural Census.



Between 1960 and 1970, the Agricultural Census divided sheep into these six categories.



Conclusion: No changes to data model

- Thematic expansion required addition of nCube and Rate Groups
  - But even this required no change to table structure
- Current system still mono-lingual, but minimal change needed to make it multi-lingual
  - Almost all text held in g_data_ent, so only change needed is adding a child table to this
- Biggest actual change was developing a PL/SQL function that populates the data map from nCube definitions
  - Enables g_data_map and g_data_coord to be treated as black box
  - Function developed for Oracle but ported to Postgres

Conclusion: Need for standard classification

- Need to span censuses means “Collections” and “Studies” not helpful
- Ten “themes” defined based on exhaustive analysis of chapter titles in statistical atlases
- Need to replace Variable Groups/nCube Groups/Rate Groups by sub-Themes
- Need a standard classification
  - Even DDI3 does not define one
  - Allows incorporation of existing classifications such as ISCO, but higher level framework needed
  - Existing thesaurii not well suited

Conclusion: Universes really matter

- Unproblematic in conventional study
  - “Universe” = what an nCube adds up to
  - Universe = group covered by the survey
- Much more complex when merging many surveys of different types
  - nCube combining variables for gender and age may cover:
    - All persons (Census of Population)
    - All deaths (Data from Registrar General’s reports)
    - All sheep! (Farm Census)
- Universes and measurement units defined as entities, not just attributes
  - Mappings to create derived values check that nCubes share universes as well as being based on variables that are mapped to one another

Conclusion: Need better interfaces

- Biggest limitation of current system is user interfaces:
  - No searching, only browsing
  - System populated by editing scripts consisting of SQL insert statements
- Searching or a better (faceted) browser?
  - Current system allows search/browse to an admin unit, then browsing statistical data for that unit
  - Worked OK until Farm Census data added
  - Faceted browsers intuitive but computationally demanding
  - Hope to add simple free text searching of labels and text to locate nCubes of interest
- Web-based editing system should be no harder to write than existing AUO editor
  - Data models similar

Conclusion: Datasets are a bad idea

- Main justification for studying historical statistics is to understand change over time
  - Far more users want local time series than cross-sections
- Conventional organisation of data archives is large barrier to analysis
- Data Item-level DDI documentation is only part of the solution
  - “Collections” and “Studies” as barriers
  - Defining non-additive nCubes may speed metadata creation but is a barrier to analysis
- “Data Feed”: data as a web service
  - DDI needs service protocol as well as standard classification
  - ONS have rejected DDI in favour of SDMX for 2011 web service

Web sites, etc

- Vision of Britain:
  www.VisionOfBritain.org.uk
- Data Documentation System:
  www.VisionOfBritain.org.uk/data
- Great Britain Historical GIS:
  www.gbhgis.org
  www.port.ac.uk/research/gbhgis
- Mailing list:
  www.jiscmail.ac.uk/lists/gbhgis

Great Britain Historical GIS Project:
A Vision of Britain though Time

University of Portsmouth

"Creating DDI Metadata for a Diverse Body of Aggregate Statistics: Experience with the GB Historical GIS"

Humphrey Southall (University of Portsmouth)

The Great Britain Historical GIS combines boundary maps for the main British statistical reporting units with a wide range of demographic, economic and social statistics from the last 200 years. Although the original system held statistics in many separate database tables, a new architecture was developed in 2001-4 holding a large subset of the statistics collection in just one column of a single large database table, facilitating graphical presentation via the public and very popular web site "A Vision of Britain through Time". In this architecture, the meaning of each data value – what it measures, as distinct from where and when – is recorded via a metadata framework based fairly closely on DDI2. This paper focuses not on the overall architecture but on the decisions and trade-offs involved in holding a diverse library of statistics in this single data structure. The original system was mainly focused on census statistics and vital registration statistics. More recent additions include labour market data, a large body of statistics taken from British Agricultural Censuses between 1867 and 1971, and election results: the number of votes for each party in each constituency in each British parliamentary election since 1832. Almost no architectural changes have been needed to accommodate these extensions.