

8 Great Britain

Microdata from the 1851 and 1881 Censuses

Matthew Woollard

Introduction

The 1851 census of Great Britain was the fifth decennial census. It was the second of the “modern” censuses which collected individual-level information on the whole of the population, but was also a considerable extension on its predecessor. The 1841 census collected nominal information on sex, age (rounded down to the nearest five), occupation and place of birth (whether born in the county of residence or elsewhere). In 1851, a number of refinements were made. Additional information was collected on relationship to head of household and marital status. More detailed information on place of birth was also requested as was the exact age of each individual. Finally, a question on infirmity was included. It is this additional information which makes the 1851 British census the earliest which has been considered useful for demographic and social and economic research in the UK. It is perhaps unfortunate that the earlier census of 1841 has received considerably less attention than its successor.

Matthew Woollard is a Senior Research Officer on the 1881 British Census Project in the Department of History at the University of Essex. He is the author of a number of articles on historical computing and is currently researching the evolution of occupational classification schemes in Britain, 1672-1921.

For more information about the historical censuses of Great Britain, see:
<http://www.essex.ac.uk/history/hcssrg/>.

The 1881 census was the ninth census of Great Britain, though it must be noted that between 1861 and 1891 the censuses of Scotland were administratively separate from those of England and Wales, when separate Acts of Parliament were enacted. (Drake 1972). From 1901 the scope of the Census Act was for Great Britain. To all intents and purposes, however, the same information was collected in 1881 and contemporaneously processed in a very similar manner.¹

Creation of the Dataset

Original Collection of the Material

Each of the British censuses in the years from 1841 to 1901 collected information in the same general manner. Enumerators left printed schedules for householders to complete. These were collected in due course by the enumerator, checked (often perfunctorily), copied out into what has become known as a census enumerators' book (CEB), and then forwarded, via (usually two) registrars to a temporary census office, where tabulations of the information were made. It is perhaps interesting to note that from 1841 onwards a Welsh translation of the schedule was provided, but these schedules, if completed in Welsh, had to be translated into English by the enumerator into their enumeration book. For further details, see Mills and Schürer (1996b) and Higgs (1996). The CEBs are the basis for all machine-readable representations of the census, as all schedules were supposed to be destroyed subsequent to analysis of the data found in the CEBs.

Creation of 1851 Database

There is not space here to do justice to the process by which the 1851 sample census was created. For full details the reader is pointed to the end of project report submitted to the

¹ In Scotland in 1881 a question was asked about the number of rooms with one or more windows. This question was not repeated in England. An additional question was inserted onto the schedule in red ink after the schedule had been printed which read "In column 6...the word "Gaelic" to be added opposite the name of each person who speaks Gaelic habitually."

Social Science Research Council (now known as the Economic and Social Research Council, Anderson *et al.* 1980) and to the documentation provided with the dataset (Anderson 1987). Details are perhaps most easily available in Anderson *et al.* 1977.

Unlike the 1881 British census database, the 1851 dataset is a two per cent stratified cluster sample, a form chosen to maximise representativity. The sampling process was as follows: all settlements were stratified into four groups dependent on population size. A fifth group of institutions was also included. The main groups, those of towns, large non-urban settlements and other places were represented by every fiftieth CEB. The fourth and fifth strata were selected differently: small non-urban settlements were selected by taking the whole of every fiftieth place in the census reports, which allows communities to be studied *in extenso* rather than in a fragmentary fashion, as would have happened otherwise. The fifth group, institutions, were sampled by “treating all institutional populations as if they were a continuous list and then systematically selecting twenty individuals from each successive one thousand names.” (Anderson 1987). The documentation notes most of the problems associated with this form of sampling, and urges users to beware that this is a national sample, and, that, with the exception of small non-urban communities may not be representative at any lower level. A “group” of data is described as a cluster; there are a total of 980 clusters, each representing a geographical or institutional division.

The whole 1851 dataset was keyed in exactly as in the original CEBs, with additional fields allowed for comments by data-inputters. The documentation provided with the dataset (Anderson 1987) provides a full guide to the decisions made during data entry. As Table 8—1 shows, the dataset as currently distributed also contains a number of coded variables.

The most important of these, OCCCODE, is not provided within the original data tables as supplied by the HDS. The file containing these codes will be supplied by the HDS and will need to be integrated into the data. The code used by Anderson and his team is one of the most detailed and useful of all classification schemes used by historical researchers. The code is a 57 digit code which attempts to use all the information found in the occupational string. The exact sequencing of the 16

Table 8—1. Variables Available for 1851 Sample of the Census of Great Britain

Variable Name	Definition
CLUSCODE	Cluster code
NUMOFIND	Number of individuals in the household
NEWSHSIND	Cluster code number or linecode (Indicating the way in which the enumerator used the schedule number)
SCHELDNUM	Schedule number
NEIGHINF	Information about neighbouring houses, usually indicative of whether the neighbouring house was uninhabited or being built
PLACERES	Street or place name
ADDRESSD	House number or house name
OTHERINF	Comments
OTHERINFA	Where OTHERINF exceeds 80 characters
OTHERINFB	Where OTHERINF exceeds 160 characters
Field11	In the dataset as distributed there are two "fields" which are always null. Their use is unknown
Field12	
CLUSLOCN	Name of the cluster
CNTYCLUS	Name of the county in which the cluster is located
GRIDCLUS	Grid reference of cluster
GRIDCODE	Grid reference inference code
CNTCLCOD	County code
PLCTCODE	Place type code
SEQOFIND	Sequence of person in household
PRENAME	First name
SURNAME	Surname
RELTOHED	Relationship to head of household
MARRSTAT	Marital status
AGEIFMAL	Age of males
AGEIFFEM	Age of females
CNTYBRTH	County of birth
OCCUPATION	Occupation
DITTOA	Additional occupational information
DITTOB	Additional occupational information

Table 8—1. Variables Available for 1851 Sample of the Census of Great Britain (Continued)

Variable Name	Definition
CNTRYBRTH	Country of birth (where recorded)
COMMBRTH	Community of birth
DISABLED	Infirmity (as recorded)
RELNCODE	Code for relationship to head of household
CNTYCODE	Code for country of birth
CTRYCODE	Code for county of birth
COMMCODE	Community code
OCCCODE	Occupational code

component parts of this code is shown in Table 8—2. Unfortunately not all occupations are classified, and there are some uncorrected errors within them. For example the first six digits represent a unique occupational code, of which the first four digits comprise the industrial sectors and sub-sectors.

“Relationship to head of household” is also coded (RELNCODE) and is complete for all records.

The complete 1851 census of Great Britain dataset contains 398,401 person records, representing approximately 82,645 households.

Creation of the 1881 Database

Census Enumerators’ Books are also the main source for the 1881 database. In 1987 the Genealogical Society of Utah (GSU), along with the Federation of Family History Societies, initiated a project to make a machine-readable transcription of the whole of the 1881 census for England and Wales. Subsequently a similar project was started for Scotland. The fruits of this remarkable data collection process is publicly available on CD-ROM (GSU 1999). The basic aim of the GSU was not, of course, to provide a service for academic research, but to assist family historians in their research, and this has meant that while the data collection for this particular project was carried out with the assistance of academic researchers, not all of the original document was transcribed and made machine-readable.

Table 8—2. The 1851 Occupation Code Components

Variable Name	# of Cols.	Variable Description
OCCID1	6	Basic occupational title
OCCUINF	1	Inference code relating to family occupation
OCCUECAC	2	Economic activity
OCCUPATT	2	Code for personal resources brought to job market
OCCUHIER	2	Code summarizing an individuals position in a hierarchy of control
OCCUPART	1	Number of partners with whom the individual trades
OCCUNOSM	4	Number of adult males employed
OCCUNOSW	4	Number of adult females employed
OCCUNOSB	4	Number of boys employed
OCCUNOSG	4	Number of girls employed
OCCUNOSU	4	Number of employers where sex /age not specified
OCCUACRT	5	Total number of acres farmed
OCCUACRN	5	Total number of marginal acres farmed
OCCUNOCC	1	Number of occupations individual is employed in
OCCUID2	6	Second basic occupational title
OCCUID3	6	Third basic occupational title

In most cases, the information lost is peripheral, but in one instance this may pose serious problems.

The remainder of this section will describe briefly the process by which the database was created. Further details can be found in Young (1994) and Woollard (1996). The early planning phases did not include Scotland since it was not initially included in the project. Once it had been decided that the whole of the 1881 census for England and Wales was to be made machine-readable, each extant CEB was microfilmed. Hard copies were made of each microfilm and broken down into manageable batches. A total of just over 45,000 batches were made which roughly equates to the number of CEBs. Each batch was transcribed twice, mainly by members of family history

societies, after which a checker compared the two transcriptions to eliminate errors or make judgements on illegible entries. The batches were then forwarded to a management centre, where further evaluative procedures were carried out. From the management centre, the corrected transcriptions were then sent to a number of data entry centres, where the transcriptions were made machine-readable. (It was originally planned that every batch would be data entered twice, one keyboard operator typing over the previously entered data. This may have occurred in many instances, but there are obvious indications within the database that data entry only occurred once. For example, it is extremely unlikely that two data entry operators would mis-key the same incorrect letter in a word).

The computer program used by the GSU automatically added some information to the data. Once input, the transcription sheets, along with the machine-readable version were returned to the project management centre, an audit was performed, usually by checking the quality of the data entry. Once the audit was complete, the machine-readable data were sent to Salt Lake City, where indexing and formatting took place.

It is important to keep in mind that the original aim of the project was not to create a database of the 1881 census, but to create a microfiche index. The work was created for genealogists rather than historians, and the importance of keeping the whole of the source in a format as near to its original as possible was not always considered necessary by both the creators and those involved in the production of the database. This means that integral parts of the original CEBs may have been lost—it is not always clear whether the indexing procedure has always kept the records within the CEBs in their original order (though it should be possible to reconstruct this in those places where it has not been kept).

The data were, after production of the microfiche, supplied to the History Data Service (HDS) at the University of Essex for distribution. Data were made available to the HDS on a county-by-county basis and it was initially distributed on such a basis. Unfortunately, the dataset as supplied by the GSU to the HDS was not completely readable from the tapes and re-supply of the magnetic tapes was not possible. In 1999 the GSU published the whole index on CD-ROM (GSU, 1999), which

allowed researchers in the History Department at the University of Essex to extract the whole dataset. It is this version which has been deposited and is presently distributed by the HDS. It differs markedly from the original transcriptions supplied by the GSU and has been subject to a certain amount of further cleaning and verification, though, at the time of writing this has not been completed. The version currently available thus neither represents the whole of the data in the original CEBs nor the whole of the material collected by the GSU. Furthermore, at the time of writing only that part of this “second-generation” dataset which relates to England, Wales, the Channel Islands and the Isle of Man is available.

Because of the manner in which the GSU published their data, it is still impossible to be certain of the number of person records in the 1881 dataset. This is because records relating to uninhabited houses and houses in the process of being built were given records in the same format as a person record. These have almost been identified totally. Table 8-3 gives an indication of the number of person records available for each county in England and Wales, along with the published populations of those counties. It should also be noted that the administrative geography of England and Wales was notoriously complex and the figures reported may not exactly relate to the same geographical unit. At this time, a similar exercise has not been carried out for Scotland.

Electronic Formats

The version of the 1851 dataset supplied by the History Data Service, where the data have been deposited, is in ASCII quote-separated text. Each of the 980 cluster files are provided both separately and concatenated into ten larger files. A single flat file dataset, augmented by the History Department at the University of Essex has also been deposited at the HDS with some enhancements, adjustments and minor additional cleaning. This new version will hopefully be made available in the autumn of 2000.

Presently, the 1881 data is distributed in a flat-file ASCII version, by application to the History Data Service, UK Data Archive, University of Essex. Contact email: hds@essex.ac.uk.

Table 8–3. Comparison of Population from the Published 1881 Census Report with the Records in the GSU Created Data.

Registration county	Published Population	Dataset Population
Middlesex	2,920,485	2,914,058
Surrey	1,436,899	1,418,998
Kent	977,706	995,089
Sussex	490,505	491,476
Hampshire	593,470	598,009
Berkshire	218,363	218,836
Hertfordshire	203,069	200,753
Buckinghamshire	176,323	176,274
Oxfordshire	179,559	179,958
Northamptonshire	272,555	274,162
Huntingdonshire	59,491	57,815
Bedfordshire	149,473	151,014
Cambridgeshire	185,594	184,593
Essex	576,434	575,329
Suffolk	356,893	354,613
Norfolk	444,749	448,031
Wiltshire	258,965	257,669
Dorsetshire	191,028	191,307
Devonshire	603,595	606,334
Cornwall	330,686	330,182
Somersetshire	469,109	468,475
Gloucestershire	572,433	570,711
Herefordshire	121,062	119,367
Shropshire	248,014	251,734
Staffordshire	981,013	983,001
Worcestershire	380,283	380,302
Warwickshire	737,339	734,749
Leicestershire	321,258	322,983
Rutlandshire	21,434	21,415
Lincolnshire	469,919	466,381
Nottinghamshire	391,815	393,120
Derbyshire	461,914	456,527
Cheshire	644,037	643,714
Lancashire	3,454,441	3,452,646

Table 8–3. Comparison of Population from the Published 1881 Census Report with the Records in the GSU Created Data (Continued)

Registration county	Published Population	Dataset Population
Yorkshire	2,886,564	2,892,563
Durham	867,258	866,451
Northumberland	434,086	433,869
Cumberland	250,647	250,659
Westmorland	64,191	64,089
Monmouthshire	211,267	210,612
Glamorgan	511,433	508,335
Carmarthenshire	124,864	122,911
Pembrokeshire	91,824	92,692
Cardiganshire	70,270	71,153
Brecknock	57,746	58,270
Radnorshire	23,528	23,516
Montgomeryshire	65,718	66,834
Flintshire	80,587	78,866
Denbighshire	111,740	109,864
Merionethshire	52,038	53,332
Carnarvonshire	119,349	117,826
Anglesey	51,416	51,413
Guernsey	35,257	33,702
Jersey	52,445	52,478
Isle of Man	53,558	53,472
Royal Navy	n/a	29,532
Total	26,115,699	26,132,064

Variable Availability

Table 8–1 summarizes details as supplied with the 1851 dataset. It should be noted that in the readme.txt file distributed with the data the following variables have been erroneously described: CNTYCODE should be a code for marital status; CTRYCODE should be CNTYCODE (i.e. code for the county of birth) and finally the field entitled COMMCODE should be

CTRYCODE. Not all of either the values or implied values for these variables have been coded. For example, someone whose place of birth was recorded as Dublin will have a value for the county of Dublin but not a country code for Ireland unless that is explicitly stated in the country of birth field. With patience these codes can be augmented.

Variable availability for the 1881 dataset has changed over time. Data extracted from the GSU CDs and parsed locally are not perfect in all respects. Work is ongoing to reshape the dataset and availability may alter. Variables presently available include those in Table 8-4. As modifications are made and derived variables are added this list will need augmenting, but at present no further variables can be guaranteed. The table also shows the provenance of each variable. Those given as “data” are those taken from the original CEBs; those described as “derived” have been created from data within the dataset. Those that are described as “data (parsed)” are those which have been parsed from the GSU published CD-ROMs into their original constituent parts, though, in the version currently available there are potentially some errors because of the inconsistent manner in which the GSU merged the original entries.

Availability of Data

The 1851 and 1881 datasets are available only to academic researchers in the United Kingdom. Contact the History Data Service at the United Kingdom Data Archive, University of Essex, C04 3SQ, UK. The web site is: <http://hds.essex.ac.uk/>.

Research Possibilities

Despite its general availability, the 1851 census microdata sample has not been used widely. This may be a consequence of only data being coded for a one-fortieth sub-sample. Significant research findings on the British household and on family structure have been reported by Anderson (1988). The same author has used a smaller sample in a recent article on married women’s employment (1999).

The 1881 CEBs are one of the key sources for the study of British social, economic and demographic history. They have

**Table 8-4. Variables available in 1881
Great British Census Database**

Variable Name	Brief Description	Provenance
cdrom	Name of CD-ROM from published GSU set of 1881 CD-ROMs	Derived
ref	A composite variable, uniquely identifying each record with information from elsewhere in the dataset. The five groups of digits contain the contents of the fields proref, piece, folio, page and recid	Derived
recid	A sequential number giving a unique number relating to the CD from which the record was extracted	Derived
hid	Household identifier. Presently this relates to the cd-rom variable and was constructed by using each "head" as the first record in a household	Derived
pid	Personal identifier. This is a sequential number for each person within a household	Derived
rectype	Record type. Defines the type of record; 1 = person 2 = person in an institution (this is often wrong!) 3 = person on board a ship 4 = a non-person record	Derived
addr	Address	Data
cparish	Civil parish of enumeration	Data
ccnty	County of enumeration	Data
ccntry	Country of enumeration. [England, Wales or Scotland]	GSU added
proref	The PRO reference for each record	Data
piece	PRO piece number	Data
folio	PRO folio number	Data
page	Page number	Data
sname	Surname	Data (parsed)
pname	First name	Data (parsed)

Table 8-4. Variables available in 1881
Great British Census Database
(continued)

Variable Name	Brief Description	Provenance
mar	Marital status	Data
age	Age	Data
sex	Gender	Data
birthpl	Birthplace as extracted	Data
bparish	Birth "parish"	Data (parsed)
bcnty	Birth county	Data (parsed)
bcntry	Birth country	GSU added (parsed)
rel	Relationship to head of household	Data
handicap	Handicap	Data
occ	Occupation	Data

been widely used for local studies, as the collection of articles published in Mills and Schürer (1996a) and in the accumulative bibliography of publications using CEBs attest (Mills and Pearce). More recent articles using CEBs can be found in numerous journals, but especially in *Local Population Studies*.²

Future Plans

As described above, some minor data cleaning on the 1851 dataset has been carried out in the History Department at the University of Essex. This cleaned version has been deposited with the HDS but is not yet available for use.

In 1998-1999 a project funded by the Leverhulme Trust created occupation codes based on the 1881 census classification.³ All occupations have been assigned one of these

² Details available from LPS General Office, Department of Humanities, University of Hertfordshire, Watford Campus, Aldenham, Watford, WD2 8AT, UK.

³ See Woollard 1999 for further details.

three-digit codes, but have yet to be integrated into the dataset. The same project also provided funding for the allocation of standardised birth places on a parish level.⁴ Place name standardization was largely completed on the originally deposited data but because this amounted to less than 75 percent of the data, the project is incomplete. Also, our re-creation of the dataset from the GSU CD-ROMs—and the GSU additions and inclusions of country of origin which were not in the original CEBs—has led to some problems. Resolution of these problems has led to some heroic assumptions which can be resolved only when the data have been fully checked. Marital status and relationship for all records have been classified under the auspices of the ESRC-funded Future of Work Programme. All these classifications will be encoded to the data for eventual release.

A further tool for analysis of the 1881 CEBs is a parish-based geographic information system (GIS). Parishes are the smallest geographic area for which data can be extracted on a national level. The Leverhulme Trust funded project at the University of Essex commissioned such a map from the Historical GIS team (see Southall and Gregory, Chapter 19 in this volume).

Bibliography

- Anderson, M., B. Collins and C. Stott (1977) 'The national sample from the 1851 census of Great Britain—sample and data handling,' *Urban History Newsletter* 1977, 55–59.
- Anderson, M., B. Collins and J. C. Stott (1980) 'Preparation and analysis of a machine-readable national sample from the enumerators books of the 1851 census of Great Britain.' End of Grant Report, January 1980. Copy filed in the British Library. (A photocopy is held in the Cambridge Group library.)
- Anderson, M. (1987), 'National sample from the 1851 census of Great Britain. Introductory user guide.' Department of Economic and Social History, University of Edinburgh.
- _____. (1988) 'Households, families and individuals: some preliminary results from the national sample from the 1851 census of Great Britain,' *Continuity and Change* 3, 421–438.

⁴ See <http://www.essex.ac.uk/history/hcssrg/levfinal.htm> for further information on this project.

- _____. (1999) 'What can the mid-Victorian censuses tell us about variations in married women's employment?' *Local Population Studies*, 62, 9–30.
- Drake, M. (1972) 'The census, 1801–1891,' in E. A. Wrigley (ed.) *Nineteenth-century society. Essays in the use of quantitative methods for the study of social data*. Cambridge: CUP: 7–46.
- Genealogical Society of Utah (1999) *1881 British census and national index*. [25 CDs]. Salt Lake City, Utah: GSU.
- Genealogical Society of Utah, Federation of Family Historians, (1997) *1881 Census for Great Britain* [computer file]. Colchester, Essex: The Data Archive [distributor], 29 July 1997. SN 3643.
- Higgs, E. (1996) *A clearer sense of the census. The Victorian censuses and historical research*. London: HMSO.
- Interdepartmental Committee on Social and Economic Research (1951) *Guides to the official sources. No. 2. Census reports of Great Britain, 1801–1931*. London: HMSO.
- Mills, D. R. and C. Pearce (1986) *People and places in the Victorian census. A review and bibliography of publications based substantially on the manuscript census enumerators' books, 1841–1911*. Cheltenham: Historical Geography Research Series, 23.
- Mills, D. R. and K. Schürer (1996a) (eds) *Local communities in the Victorian census enumerators' books*. Oxford: Leopard's Head Press.
- _____. (1996b) 'The enumeration process,' in D. R. Mills and K. Schürer (eds) *Local communities in the Victorian census enumerators' books*. Oxford: Leopard's Head Press: 16–26.
- Woollard, M. (1996) 'Creating a machine-readable version of the 1881 census of England and Wales,' in C. Harvey and J. Press, *Databases in historical research*. London: Macmillan Press: 98–101.
- _____. (1999) *The classification of occupations in the 1881 census of England and Wales*. Colchester: Historical Censuses and Social Surveys Research Group Occasional Paper No. 1.
- Woollard, M. with M. Allen (1999) *1881 census enumerators' books database: description of the fields, v.0.4*. Colchester: History Data Service.
- Young, S. C. (1994) 'The British 1881 census project,' paper presented to the Association for History and Computing (UK Branch) Conference, University of Hull, 13 April 1994.

