



## Economic and Social Council

Distr.: General  
14 March 2012

Original: English

---

### Economic Commission for Europe

Conference of European Statisticians

Group of Experts on Population and Housing Censuses

Fourteenth Meeting

Geneva, 24-25 May 2012

Item 3 of the provisional agenda

Access to microdata

### Plans for Access to United Kingdom microdata from 2011 Census

Note by Office for National Statistics, United Kingdom

#### *Summary*

This paper outlines the background to the design of the proposed United Kingdom (UK) microdata products to be developed following the 2011 Census in England and Wales, Scotland and Northern Ireland, including an overview of what was made available in 1991 and 2001.

The impact of relevant changes in UK legislation and statistical disclosure control methodology between 2001 and 2011 is explained, as are the activities around user consultation and design of microdata specifications.

Plans to date for 2011 Census microdata products and corresponding levels of access are then described.

## I. Introduction

1. This document has been prepared by Emma White with support of Denise McGregor, Keith Spicer and Joe Traynor of the Office for National Statistics and Keith Dugmore of the Demographic Decisions Ltd.

2. The census is the most complete source of information about the population in the United Kingdom, and has been held every 10 years since 1801 (with the exception of 1941, while a 10 per cent sample census was held in 1966). The range and scope of census data are wide and detailed, with the information provided being used for planning and resource allocation, as well for academic, commercial and other purposes. A new dimension to this data collection activity is the European requirement for member states to provide datasets of information on their populations to Eurostat by 31 March 2014.

3. The most recent Census Day in England, Wales, Scotland and Northern Ireland was 27 March 2011. The Office for National Statistics (ONS) has responsibility for the delivery of the England and Wales census results, while the National Records of Scotland (NRS) and Northern Ireland Statistics and Research Agency (NISRA) have responsibility for delivering results for Scotland and Northern Ireland respectively. Following the separate release of national information from the three offices from July 2012, ONS will coordinate UK data releases.

4. ONS is confident that the target national response rate of 94 per cent for 2011 Census in England and Wales will be met, as will the target of a response rate of over 80 per cent in every local authority (LA), while the voluntary Census Coverage Survey had a response rate of approximately 90 per cent. Meeting these targets means that ONS is on track to deliver England and Wales population estimates within an error of +/- 0.2 per cent, and all LA population estimates within +/- 3 per cent. Early assessments also look promising for meeting the item non-response target of no more than 5 per cent missing responses for 70 per cent of questions.

5. An independent customer survey has concluded that the methods adopted for coverage assessment, coverage adjustment, and quality assurance of the 2011 Census in England and Wales “give confidence that the resulting final census population estimates will be better than any other method and will be suitable for use in resource allocation and planning”.

6. To date, more than 100 new tables have been designed for 2011 Census following user consultation, an increase of 25 per cent on 2001. Some present data collected for the first time in 2011, others bridge user-identified gaps in 2001 products. If the data quality permits, information will also be made available on alternative population bases such as workplace, workday, student out-of-term, short-term migrant, and majority-of-time. These plans widen the census product base, provide better access to census data, meet the needs of more users, and utilise census data more effectively, as well as delivering added value for the investment in the census.

7. Consultation has highlighted the importance of making detailed, accurate data available for small population groups. A suite of products designed around these groups will therefore be produced. Furthermore, as ethnic and religious groups that do not have tick-boxes on the census questionnaire have made it clear that they want to be recognised at the same time as those who do, the second release of census tables (the first in which ethnic and religious group data are available) will include an estimate of all ethnic and religious groups for whom ONS has a classification code, down to the lowest level of census geography, output area (subject to statistical confidentiality constraints). As with alternative

population base estimates, there are no existing United Kingdom sources equivalent to the census for these data. The same is true for microdata.

## II. Microdata in United Kingdom

8. The ONS Longitudinal Study (LS) collects information on one per cent of the population of England and Wales. The LS currently holds individual level data for more than 500,000 people, including data from all censuses since 1971. Microdata have previously been produced by anonymising LS data through the removal of names and addresses. While the sample within this dataset is anonymised, they are still individual level data that have not been disguised or aggregated. Accordingly, they are subject to controlled access arrangements.

9. Samples of anonymised records (SARs) were first made available based on the 1991 Census. Two products were released – each for Great Britain and Northern Ireland separately. Details are in Table 1.

Table 1  
1991 SARs

	<i>Sample</i>	<i>Geography</i>	<i>Licence</i>	<i>Availability</i>	<i>Cost</i>
Individual SAR	2% of individuals (>1.1 million records)	288 SAR areas with populations > 120,000	End User Licence (EUL). Application through Cathie Marsh Centre for Census and Survey Research (CCSR)	Academic users: Data downloadable in SPSS, Stata or TAB. Web based exploration also available. CCSR	Free of charge (a charge was payable for non academic users prior to October 2011)
Household SAR	1% of households (216,000 records, > 500,000 people)	Regional	As for Individual SAR	Non academic users: CD As for Individual SAR	As for Individual SAR

*Source:* Dugmore, Keith. A business case for microdata files from the 2011 Census. Final Report, April 2009. Commissioned by the University of Manchester, and carried out by Demographic Decisions Ltd, London and ONS.

10. As described in the report by Demographic Decisions Ltd, the 1991 SARs supported a surge in data-driven social science research, whereby researchers from diverse fields quickly saw the benefits inherent in the files, especially given their ease of access.

11. The 2001 Census, however, created a more complex range of products reflecting rising concerns around the disclosure of personal information. Five different products were made available via three different licensing arrangements, where more restricted access corresponded to increased levels of detail in the data. See Table 2.

Table 2  
**2001 Census Samples of Anonymised Records**

	<i>Sample</i>	<i>Geography</i>	<i>Licence</i>	<i>Availability</i>	<i>Cost</i>
Individual SAR	3% (1.84m records)	UK file (Government Office Regions (GOR), Inner and Outer London)	EUL through CCSR	Academic users: Free of charge (a Data charge was downloadable in SPSS, Stata or TAB. Web based exploration also available. CCSR	Academic users: Free of charge (a charge was payable for non academic users prior to October 2011)
Small Area Microdata (SAM)	5% (2.96m individual records)	UK file (local authorities (LA) or Northern Ireland (NI) constituencies)	As for Individual SAR	Non academic users: CD As for Individual SAR	As for Individual SAR
Special Licence Household SAR records,	1% (>200,000 records, >500,000 people)	England and Wales only (no geography)	Special Licence and Approved Researcher	UK Data Archive (UKDA) and ONS	As for Individual SAR
Individual Controlled Access Microdata Sample (CAMS)	3% (same sample as for Individual SAR)	UK file (LA or NI constituencies)	Approved Researcher	ONS Virtual Microdata Laboratory (VML)	As for Individual SAR
Household CAMS	1%	As for Individual CAMS	As for Individual CAMS	As for Individual CAMS	As for Individual SAR

12. The Approved Researcher scheme was introduced in 2007. A formal application must be submitted, and outputs from the ONS Virtual Microdata Laboratory (VML) are assessed for disclosure risk by ONS specialist staff, prior to being sent to the researcher (usually within seven days). Physical access to the VML is only via one of the three ONS offices in England and Wales. Many census users were disappointed with the access arrangements for 2001 microdata, finding that access to equivalent data was in some cases more difficult than in 1991. Nonetheless, the products have been extensively used in research and analysis, both independently and in conjunction with other datasets.

### III. Changes since 2001

13. There have been a number of changes between 2001 and 2011 that affect the ways in which ONS might approach the development of microdata products:

(a) The Statistics and Services Registration Act 2007 gives ONS legal authority to release data to Approved Researchers and criteria by which to do so.

(b) There is increased computing power both within ONS and externally. This means there are potentially more ways in which data can be provided, while users are expressing greater demand for quicker, easier access to statistics generally and microdata in particular, while also seeking more detail in the statistics that ONS provides.

(c) The UK transparency and open government agenda puts the onus on Government departments to make more information available free of charge.

(d) There are increased concerns both within Government and amongst the public about privacy and the protection of confidentiality. Requirements to protect data have increased markedly since 2001.

(e) The introduction of the Open Government Licence (OGL) allows users to copy, publish, distribute and transmit the Information (users must acknowledge source); adapt the Information; and exploit the Information commercially for example, by combining it with other Information, or by including it in their own product or application. However, users must still observe the Data Protection Act 1998, and must not misrepresent the data.

#### IV. Statistical Disclosure Control (SDC) methodology for 2011 Census

14. Following feedback on the 2001 methodology of random record swapping combined with small cell adjustment (SCA), *targeted* swapping of households on the basis of risky records will be applied to 2011 Census data across the United Kingdom. The Registrars General have agreed that small counts (0s, 1s, and 2s) can be included in publicly disseminated census tables provided that:

(a) sufficient uncertainty as to whether the small cell is a true value has been systematically created; and that

(b) such uncertainty does not significantly damage the data.

15. 'Sufficient uncertainty' is assessed by two factors; the percentage of real attribute disclosures protected by swapping and imputation, and the percentage of apparent attribute disclosures that have been created. The level of record swapping in an area is determined by the level of non-response and imputation in that swapping will be lower where more records have been imputed.

16. The effects of the agreed SDC methodology will be greater on small cell counts, but that effect dissipates at higher geographical levels. As there will be no SCA, there will be complete consistency between tables. Table 3 shows the key differences in ONS methodology between the 2001 and 2011 Censuses and their impact.

Table 3

##### Differences in SDC between 2001 and 2011

	2001	2011
Record swapping	Random	Targeted to risky records
Post-tabular	SCA in England, Wales, Northern Ireland	None
Additivity	Not in all tables	Yes
Consistency	No	Yes
SDC Harmonisation	No	Yes – though parameters may differ between different countries
Effect on small counts	Large	Some

Source: ONS SDC

#### V. SDC considerations for 2011 microdata

17. Plans for 2011 microdata will seek the best balance between risk and utility, bearing in mind:

- the balance of recoding against perturbation;

- whether samples should be mutually exclusive;
- new variables and categories; and
- licensing and access arrangements.

## **VI. Consultation**

18. The consultations on main statistical outputs for the 2011 Census has had two distinct phases, running from 14 December 2009 to 26 March 2010, and from 7 February 2011 to 28 April 2011.

19. These phases included formal consultation feedback documents for completion and return, and were supported by national public consultation events, as well as direct engagement with key users and key user groups.

20. This was followed by targeted workshops on specialist products, including microdata. Following a microdata workshop on 11 November 2011, notes, materials and questions and answers from the day were made available to census users and further feedback was requested by 19 December 2011. The documentation and consultation responses were then collated, and an external microdata expert was commissioned in January 2012 to review the submissions and provide a detailed set of microdata specifications for 2011 Census. At time of writing, this work is ongoing. The set of essential and desirable microdata specifications and the final report are due on 30 March 2012. This work has been informed by the 2011 Census Microdata Strategy Working Group (CMSWG), which was convened in May 2007, and whose membership includes representatives from ONS, NRS, NISRA, and the Centre for Census and Survey Research (CCSR) at Manchester University.

## **VII. 2011: A simpler range of products than 2001 with more straightforward access**

21. Three levels of microdata are planned for 2011 Census:

### **A. Public use file**

22. This is an innovation for ONS. The release of this non-disclosive individual-level file is in line with the UK Government's transparency agenda, and it will be available for download from the ONS website. It will be less detailed than the Individual SAR file made available through an end use license (EUL) in 2001, or alternatively could be released with more detail but no geography. These data would be easy to access and they could be used as teaching and learning instruments or for high level exploration of relationships between census variables. There is potential for this file to be used in a limited way to test syntax for errors before visiting the VML. No household file would be made available for public use.

### **B. Safeguarded file(s)**

23. This file would again be at the individual level, and would be made available to users via their desktops. It is intended to contain a similar level of detail to the 2001 small area microdata (SAM) and Individual SAR files, perhaps under similar conditions to the 2001 EUL. This file could be considered as returning to the 1991 specification for the

Individual SAR, with a bigger sample. A household file is also being considered. Work is ongoing with ONS Legal Services to define 'safeguarded' and consider suitable licensing arrangements.

### C. VML files

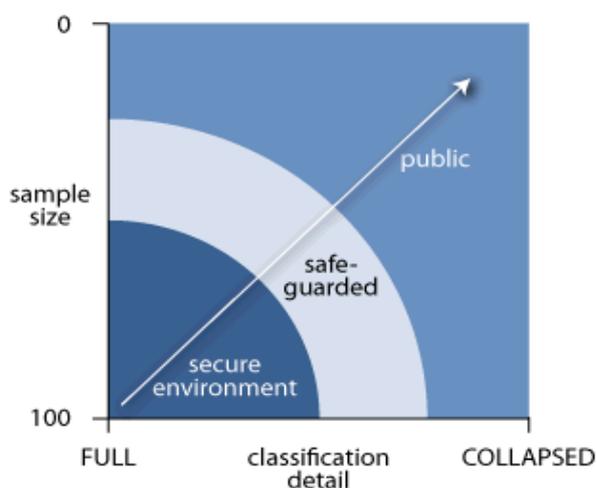
24. These anonymised but identifiable data will be held in a secure environment and will be generally similar to the 2001 controlled access microdata sample (CAMS), both at the household and individual level, and are likely to contain a sample size of between three and five per cent.

Figure 1 below summarises the levels of access for 2011.

Table 4 describes the proposed 2011 microdata products.

Figure 1

**Diagram of proposed access levels: sample size (per cent) against level of detail**



Source: ONS Data Visualisation

Table 4

#### Proposed 2011 microdata products

<i>File</i>	<i>Detail</i>	<i>Geography</i>	<i>Licence</i>	<i>Availability</i>
Public Use SAR	3%	GOR and less detail or no geography and more detail	OGL	Download
Individual SAR	5%	LA - same level of detail as 2001 Individual SAR	Safeguarded	Special User Agreement Download
Household SAR	1%	GOR or none	As for Individual SAR	As for Individual SAR
Individual CAMS	More detail than 2001? Possibility for increased sample size	At least the same level of detail as in 2001	Approved Researcher	ONS VML (Secure environment)
Household CAMS	As for Individual CAMS	As for Individual CAMS	As for Individual CAMS	As for Individual CAMS

Source: ONS

## **VIII. Timing of release**

25. Under current plans, the 2011 Census microdata files would be made available during 2013. User feedback indicates that the order of preference for release is Individual and Household SAR (Safeguarded), then the Individual and Household CAMS (VML) and finally the Public Use SAR. Users have also requested that they are kept informed of progress on the development of these products and timing of their release.

## **IX. Web-based dynamic prospectus**

26. ONS's 2011 Census prospectus has been designed as an online dynamic document which will be first published in March 2012. Users can sign up so that updates will be emailed to them as they happen. Microdata files are a key census product and are treated as such in the prospectus.

## **X. User advice, training and support for products**

27. Planning is underway to ensure that suitably trained staff are in place to provide advice, training and support for 2011 Census microdata products. ONS will liaise with CMSWG and the Economic and Social Research Council (ESRC) to consider how best to take this forward, given that ESRC, who have previously funded census services including CCSR, will be launching a new national data service following the expiry of their current round of census funding in July 2012.

## **XI. Conclusion**

28. ONS is aware of the importance of SARs for users – features include the larger sample size, the opportunity for exploratory analysis on rich and flexible datasets, the opportunity to test outputs before requesting commissioned tables, and the ability to perform multivariate statistical analysis, all using records that are comparable with the standard census outputs, since they have the same source.

29. The United Kingdom is now in a position where powerful time series analyses can be performed on microdata records, as there are files available from the 1991 and 2001 Censuses, while files for 2011 will be made available in 2013. Furthermore, the production of microdata files for census years prior to 1991 is under consideration.

30. The 2011 Census product set has been developed based on user-centred design principles. These principles also apply to microdata, bearing in mind the added complexity for SDC, licensing and access resulting from the level of detail in the dataset. Nonetheless, ONS is confident that fit-for-purpose, appropriately accessible UK microdata files will be made available in 2013.

## **References**

Dugmore, Keith. A business case for microdata files from the 2011 Census. Final Report, April 2009. Commissioned by the University of Manchester, and carried out by Demographic Decisions Ltd, London.

---