

Analytical use of 2011 Census data

Issue

1. Identifying potential analytical uses of the 2011 Census microdata, to inform census policy and communication.

Action sought

2. CGAG members are invited to:
 - comment on principles for the use of census data (para 3), and specifically on access to census microdata in ONS or elsewhere (para 4);
 - indicate the kind of uses they have in mind for 2011 Census data (para 5),
 - and specifically indicate the extent to which their departments are likely to want to use 2011 Census data as a sampling frame for follow-up surveys.

Principles for the use of census data

3. Drawing on previous censuses, and bearing the Hannigan ethos in mind, ONS is drawing up a set of principles to increase the value of 2011 Census data while maintaining its integrity and security. The following principles are, we hope, uncontentious:
 - Planned uses of census data should add to public reassurance that their data are being used beneficially and safely. Use of the data should not have a negative impact on public confidence or response rates, as far as we can assess this.
 - Personal census data should be used for a variety of analytical purposes enabled by the Statistics and Registration Services Act (SRSA), including linkage with other ONS and GSS data. While data matching and other analyses for statistical purposes are legal, the SRSA does not provide a legal framework for data sharing for any non-statistical purpose.
 - Detailed proposals for analytical access to census microdata will be developed as part of a wider, research data access

strategy. This will be primarily for ONS data but will need to recognise developments beyond ONS, not only in other departments and the devolved administrations, but also the ESRC's research data strategy.

- As for previous censuses, and in line with the census White Paper, Parliament will be informed of all potential uses of census data in advance of the census, and information will be provided as part of the census publicity strategy where possible.
- Responsibility for putting these principles into practice in ONS will rest with the Census Information Asset Owner, who is then responsible for ensuring that the asset is fully used for the public good, that access is monitored, and that risks to compiling and maintaining the asset are managed.
- That we seek, via the UK Census Committee, a harmonised UK approach to analytical use of census data.

4. Still under discussion is the issue of where to hold census microdata, in order to maximise 2011 census response and uphold the commitment to 'no data sharing'. ONS is strongly minded to present two further principles, subject to agreement that these can be operationalised to the benefit of the GSS:

- No identifiable personal data will be passed outside ONS for analytical uses, except where required to Service Providers and Approved Researchers as defined in the Statistics and Registration Services Act.
- However, data from other government departments can be linked to census data, within ONS, for statistical purposes. This can be for both the statistical purposes of ONS and the GSS in other departments, but not for any non-statistical purpose. Identifiable data will not be passed to another part of the GSS for them to link themselves.

Proposed uses of census micro data

5. Proposed uses of census micro data fall into four main categories (for further detail see Annex):

- Production of outputs either by ONS (e.g. thematic analysis of census data for wider dissemination), or by other analysts using de-identified datasets (e.g. Sample of Anonymised Records (SARs)). This potentially includes outputs derived from UK level outputs/hypercubes, etc, although this needs further study, including on the extent to which we need to, and would be able to, share identifiable person level data

between the three UK Census Offices (NISRA, GROS, ONS)

- Linkage with other ONS/GSS data sources to produce ONS statistical outputs
 - Longitudinal Study,
 - linkage with ONS Annual Survey of Hours and Earnings (ASHE),
 - linkage with NHSCR mortality data for analysis of life expectancy by population sub-groups,
 - linkage with DCSF and DWP data to improve population estimates
- Use of census data as a sampling frame for further follow-up research commissioned by both ONS and other government departments
- Quality assurance of existing survey statistics (e.g. non-response studies linking census characteristics to survey non-response data)

The Legal Framework

6. The Statistics and Registration Services Act (SRSA) has changed the legal context surrounding the use of census data and has the potential to hugely increase its analytical value. Sections 38 and 26 give the legal power to use census data for the purposes of the Statistics Board, including linkage of data. The Annex provides more detailed discussion of the legal context in relation to the proposed analyses.

Discussion

Passing identifiable data outside ONS (sharing data)

7. Although the new legal framework enables all of the analyses listed above, from a statistics policy point of view we would not want planned uses of census data to have a negative impact on public confidence or census response rates. It is unlikely that a policy which allowed us to pass identifiable data outside ONS would be considered a fair use of census data. This would undermine the census publicity strategy and messages on confidentiality, planning for which currently assumes there will be no data sharing.
8. If we do confirm a strategy based around 'no data sharing', the implication is that no identifiable data would be allowed outside ONS except for service providers processing census data, eg under the LMUK contract, or with the NHS IC for the Census LS link. ONS would have to carry out data matching on behalf of other GSS departments, and all analytical access for approved researchers will

need to be facilitated on-site at ONS. There is no clear evidence on the impact on census response (indeed we know from anecdotal evidence that many people already assume census data is widely shared across government). However, any negative media coverage and subsequent reduction in response would harm both the quality of the census itself and therefore the outputs considered here. If there is a strong case for a different approach, and ONS agreed to it, then the 2011 Census publicity strategy would need to reflect that 'ONS will hold data securely' rather than that 'there will be no data sharing'. Some data matching proposals will involve data sharing, albeit with ONS as the data controller giving access to the data.

9. In line with the Census White paper and the National Statistics Protocol on Data Matching, Parliament should be informed of proposed uses of census data prior to the census itself, and details included in census publicity where possible (e.g. follow-up surveys were mentioned on the census information leaflet included with the census form in 1991).

Providing access to linked census data

10. The datasets created for the analyses outlined above are also likely to have secondary uses, and requests to use the data or for new proposals for analysis from researchers outside ONS are likely. The approved researcher section of the SRSA provides one way of allowing this, and we already have in place the Microdata Release Panel, enhanced by an extra layer of controls where needed e.g. the LS Research Board. Further work is needed to develop a policy for approval of future proposals for analyses using personal census data and protocols for data usage (e.g. when and where data can be held, by whom and for how long). It would be inadvisable to take too many decisions affecting data access in 2009 given the speed at which technology develops, other than to agree the policy and principles. If we do this we may make decisions which limit our usage in the years before 2021. Nevertheless, we clearly need to address this within ONS's research data access strategy, ensuring that the basis of access to personal census data, and approval of future analysis proposals, are agreed and communicated.

England and Wales or UK wide policy?

11. A UK consensus on analytical use of census data has not yet been agreed but would be beneficial. This would enable GSS analytical outputs to be based on UK rather than England and Wales data in more instances, and a consistent UK policy would be easier to present to the public.

Paul Allin

Deputy Director: Household, Labour Market and Social Wellbeing, ONS

26 June 2009

Annex

Proposed uses of census micro data

12. The following uses of census micro data have already been identified in addition to the production of the population estimates (these mostly build on similar developments in 2001, although four bullets marked * are, we believe, new):

- **Series of thematic ONS analytical outputs** - Following the 2001 Census, ONS produced the thematic 'Focus On' series of publications drawing heavily on census data for a number of these. 2011 Census data has great potential to inform analysis of ONS priority themes - ageing, children, social wellbeing and globalization. ONS will be considering options for dissemination of 2011 data against these and any other emerging priority areas, in consultation with stakeholders. This will include exploring the need for UK level micro datasets and outputs.
- **Sample of Anonymised Records (SARS) / Controlled Access Microdata Sample (CAMS) and similar de-identified datasets created for analysis.** As for 2001, plans are underway to develop and release de-identified census records and to have controlled access to more detailed microdata where required. A consultation of user needs has been carried out and findings will be published at the end of next month (October 2008), but it is likely there will be demand for a higher percentage of records to be included in the sample (in 2001 it was 3%). Statistics Canada now allow approved researchers access to 100% de-identified census records, Australia 20%, and France has just passed new legislation to allow greater access to French census records.
- **Follow-up surveys** - there is some interest in using census data as a sample frame to conduct follow-up surveys on topics/small population sub-groups that cannot be covered adequately in the census itself. The window for this is relatively tight, because census data will be increasingly out of date and they will anyway not be available until some months after census day. If departments want to explore this, both they and us will need to consider this over the next few months, in order to allow time for the planning of the survey and the availability of the sampling frame.

There are precedents. Personal census data have been used as a sampling frame to identify samples of small socio-economic groups which could not be effectively identified using other means. This includes follow-up surveys after the 1991 Census on the workforce in professions allied to medicine, qualified social workers and probation officers, surveys of households in privately rented properties, and surveys of vacant properties. In 1991, the possibility of conducting such surveys was identified in

advance in the Census White paper, and a Parliamentary Question used to confirm the surveys once these had been agreed. Similar provision for 2011 has been made in the White Paper. Interest identified so far includes surveys on Welsh language and of migrants or members of ethnic minorities.

- **Linkage with data collected under the Statistics of Trade Act / ASHE data *** - This is an ONS interest. In addition to linkage with ASHE data, linkage of census and other data collected under the Statistics of Trade Act is required for analysis of skills and economic performance data.
- **Linkage with data from other government departments ***
 - > **Demography** are putting in place secondary legislation to allow links with DCSF and DWP data for a number of purposes including quality assurance of the population estimates following the 2011 Census. Demography also plan to release linked longitudinal census and DWP Unemployment Benefit data, but following the SRSA an 'Information Order' for this linkage is required and is being prepared.
 - > **Integrated Population Statistics System** - As part of the development of a business case for an Integrated Population Statistics System it will be essential to benchmark a range of administrative data sources with the census by linking records at the individual and household level. This will provide an initial assessment of the feasibility of using administrative data to develop better sources for population statistics.
 - > **Linkage with mortality data** - the current 1% linkage of census and mortality data in the Longitudinal Study is insufficient to provide robust mortality data for important population sub-groups such as ethnic minorities and disabled people. Statistics Canada link mortality data to a 25% sample of their census records (although a lower proportion is likely to be needed for the UK because of our larger population). This linked data would also improve estimates of life expectancy and health active life expectancy.
 - > **Equalities measurement framework** - linkage of census data with that from OGDs may lead to more complete and robust statistics by ethnic group, religion, disability etc. This will further consolidate the cross-government equalities measurement framework.
- **Longitudinal Study (LS) 2011 Census link** - ONS Demography are planning to create a fifth census link by adding 2011 Census microdata to existing LS records. In addition to this, some of the linkage of census data with other government data could be carried out within the LS. That would give, for the 1% of records in the LS, longitudinal data going back to 1971 linked to the other government data. In some cases, this could

be a cost-effective way of achieving the required linkage whilst enhancing the value and scope of the LS.

- **IHS non-response study** - to improve social survey weighting and given continued concern over social survey response rates, it is likely that we will want to repeat the non-response studies conducted after 1991 and 2001 Censuses. These matched census data to survey non-responders to inform our understanding of survey non-response and allow us to improve our survey weighting.
- **Linking census data to the Postal Address File (PAF) *** - Methodology are in the very early stages of considering the benefits of linking census and PAF data to exploit census household data for improved stratification in social surveys.
- **Census address lists *** - some potential analytical requirements assume that there is a 'Census address register' that could be made available for analysis (eg spatial analyses). However, it is not yet clear whether this will be feasible, beyond any analysis needed for the census itself. This needs to recognise the data sharing agreement (between ONS, Ordnance Survey, Royal Mail and the IdEA) and that plans to share addresses with other suppliers are quite constrained. Address information collected as part of the census itself would be considered as personal census data.

The Legal Framework

13. The Statistics and Registration Act (SRSA) has changed the legal context surrounding the use of census data and has the potential to hugely increase its analytical value. However, although this new legal framework allows all of the analyses listed above, we would not want planned uses of census data to have a negative impact on public confidence or census response rates. The SRSA does not provide a legal gateway for data sharing for any non-statistical purpose.
14. Section 38 of the SRSA states that 'any information obtained by the Board in relation to the exercise of any of its functions can be used by it in relation to the exercise of any of its other functions'. The only limitation is any enactment which restricts or prohibits the pooling of data. Neither the Census Act nor the Statistics of Trade Act has any such restriction or prohibition on this department's use of the data collected under those enactments within the meaning of statistical purposes. The analytical uses above would be in line with the treatment of confidential data outlined in Section 39. Section 5 of the Census Act has always allowed potential linkage with other government datasets to quality assure the population estimates from the census data, including providing the legal basis for the data linkage within the Longitudinal Study. Section 26 of the SRSA states that the Board 'may do anything necessary or expedient for

the purpose of, or in connection with, the exercise of its functions'. This allows ONS to link census with other data to improve our analytical outputs (including our own data collected under the Statistics of Trade Act, and data from other departments e.g. mortality data).

15. Using the census data to provide these outputs would be effectively utilising our information assets within the law to maximise the public benefit in line with Hannigan III Section 6. Whilst Hannigan does not have the direct force of law, it could be argued that this and other initiatives are changing the culture in which decisions are made about utilising the value of the information assets we hold and making best use of public money.
16. Data linkage would be subject to the provisions of the National Statistics Data Matching protocol (or anything which replaces it). This requires the authorisation from the responsible statistician of the original data source and is dependent on any obligations or undertakings associated with that source. The protocol also includes provision to ensure respondents are informed of the uses to which census data will be put. The Census White paper refers to statistical use of the data including mention of linked samples and follow-up surveys, and includes provision to inform Parliament of any proposals before the census is conducted.

Relevant section from the Census White Paper

"The handling of information collected in the 2011 Census will conform entirely to the principles of fair processing of personal data and sensitive personal data demanded by the Data Protection Act 1998. This will include not only the collection and processing of data for the purposes of the Census itself, but also for any other statistical use of the data, for example as a base for creating linked samples for subsequent follow-up on topics that cannot be included in the Census itself. Any intention to use 2011 Census information for any such follow-up surveys will be announced to Parliament before the Census and made clear to the public in accordance with the ONS Code of Practice."