United Nations ECE/CES/GE.41/2010/15

Economic and Social Council

Distr.: General
22 June 2010

Original: English

Economic Commission for Europe

Conference of European Statisticians

Group of Experts on Population and Housing Censuses

Thirteenth Meeting
Geneva, 7-9 July 2010
Item 2 of the provisional agenda
Census communication and dissemination, including the use of Geographic Information Systems (GIS)

Disseminating Census information to maximise use and value

Note by the United Kingdom[1]

Summary

This paper discusses how the value of census information can be increased by extending its use. Drawing on experience in the United Kingdom, it introduces the idea of statistics being for the public good, rather than primarily for government's own purposes. The cost of disseminating information represents a very small percentage of the Census budget, and the return on investment should be maximised. Censuses create the opportunity to produce a range of statistical products, but we need to focus on existing and potential markets, and understand their different segments. A manifesto for seeking to maximise use and value can build on a range of proven options, but we also need to seek a better balance between risk and utility, guided by Jeremy Bentham's view of utility: "It is the greatest good to the greatest number of people which is the measure of right and wrong."

[1] Prepared by Demographic Decisions Ltd.

I. Introduction Statistics for the public good

1. The brief for this session includes the striking phrase "only used statistics are useful statistics", and this paper develops that theme, drawing on experiences in the United Kingdom.

2. Censuses in the four countries of the United Kingdom are carried out by three separate Census Offices: the Office for National Statistics (England & Wales), the General Register Office (Scotland), and the Northern Ireland Statistics and Research Agency. Whilst considerable efforts are made to adopt similar approaches, there are some differences in the statistics produced for each country. There are also issues concerning the provision of a single point of access to comparable statistics for all four countries – which has analogies with the wider situation across Europe as a whole.

3. Since the publication of the UK's 2001 Census, the Statistics and Registration Service Act 2007[2] has provided for the creation of a new body, the Statistics Board, with a statutory responsibility to promote and safeguard the production and publication of official statistics "that serve the public good". The Code of Practice, containing the standards against which National Statistics are being assessed, was published in January 2009[3]. Its Principle 1 is Meeting User Needs: "*The production, management and dissemination of official statistics should meet the requirements of informed decision-making by government, public services, business, researchers and the public.*" Note that this is not just the needs of government.

4. This is an encouraging starting point for users, but two other influences are also very important at the time of writing. Firstly, there is the very positive environment that has developed in recent years building on the campaign to Make Public Data Public. This encompasses many types of public sector information, ranging from statistics, maps, and government spending, to parliamentarians' expenses. The most recent development has been the launch of the website www.data.gov.uk which already offers more than 3,400 datasets.

5. However, on the negative side, the new government is committed to making the most severe budget cuts for many decades, and there are fears that the Census may not escape unscathed.

II. The cost of disseminating information, as a % of the total Census budget

6. Again quoting from the brief for this session, "Considering the huge resources normally spent conducting the census and producing the most relevant and accurate results, it is fundamental that appropriate means are identified to disseminate the census results and adequate resources are invested."

7. Whilst recently developing a business case for microdata files from the 2011 Census[4] for the University of Manchester, it was striking to look back at the budget for the 2001 Census in England and Wales. Of the total of £207 million, the bulk was spent on data

[2] http://www.opsi.gov.uk/ACTS/acts2007/ukpga_20070018_en_1
[3] http://www.statisticsauthority.gov.uk/assessment/code-of-practice/index.html
[4] http://www.ccsr.ac.uk/sars/2011/documents/businesscase.pdf

collection and processing: only £7 million (3% of the budget) being spent on outputs (Table 1).

Table 1.
2001 Census costs, England & Wales

	Cost (£millions)	*(%)*
Policy, Content & Publicity	24	12
Support Services	13	6
Geography	7	3
Data Collection	84	42
Data Capture & Processing	61	29
Downstream Processing & IT	11	5
Output Policy & Production	7	3
Total	**207**	**100**

Source: 2001 General Report

8. The 2011 Census is a major investment, costing more than £560 million for the United Kingdom as a whole. In the context of "statistics for the public good", the 2011 Census database will provide an immensely rich resource, and every effort should be made to extract the most value, given that the marginal costs of disseminating outputs are so small. To spend so much collecting and processing the data, but to then stint on creating and disseminating outputs to maximise its utility, would be a failure to achieve the best return on investment.

III. The range of potential statistical products

9. In common with other European countries, the UK's 2011 Censuses are planning to ask questions on a wide range of topics. Initially considering this from the supply side, there is scope to produce several different categories of statistical products:

(a) **Resident population: tables.** Published tables classifying the population by many different topics. These have traditionally represented much the greatest part of Census output, with Key Statistics appealing to the popular market, and more detailed cross-tabulations catering for specialist users' needs.

(b) **Resident population: area classifications.** These have proven popular in the United Kingdom, with classifications of Local Authorities being produced for several recent Censuses, and a new Output Area Classification (OAC) at neighbourhood (120 household) level being produced for the first time in 2001.[5]

(c) **Workplace population.** Tables classifying the population at their place of work by selected topics. These statistics are of especial value to retailers, who use them to evaluate sites in city centres.

(d) **Origin / Destination tables.** Commuting (journey from home address to work address); and migrants (address 1 year ago / current address). Such datasets are more complex to analyse, but academic specialists in particular have used them for important projects for government, defining travel to work areas, and analysing migration patterns.

[5] See the OAC User Group website http://areaclassification.org.uk/

(e) **Commissioned special tables.** These are ordered (at a cost of c.£150) by experienced users who require more detail than is available in the published tables.

(f) **Flexible table generation.** There are also plans for 2011 to offer users the opportunity to create their own tables for the first time, using hypercubes.

(g) **Microdata files.** These are of two types: individual Samples of Anonymised Records (SARs); and the Longitudinal Study, which links records of anonymised individuals across two or more censuses. Both these products appeal to specialist users, principally in academia.

10. The Census operation also enables the creation of two further products which are of great value to users: digital boundaries for Output Areas, and a Postcode / Output Area directory, enabling the linking of various postcoded address datasets to Census area statistics.

IV. Marketing the Census – identifying market segments

11. It is, however, essential that census dissemination is driven by an understanding of the existing and potential market. Segmenting the market, the two vital dimensions are the various sectors – types of organisations – and the degree of knowledge and time available to individual users.

A. Segmentation – Market sectors

12. The major market sectors in the United Kingdom are listed in Table 2, which also gives examples of case studies that are being used by ONS in its business case to the government for funding. Many of these users have common interests (for example, statistics being freely and easily available), but there are some differences (local authorities, for example, place less importance on consistency across the United Kingdom). Encouraged by the advent of the internet, another community – the general public – is of increasing importance. It is also necessary to recognise the role of Value Added Resellers in disseminating Census information to wider audiences (see Section 5 below).

Table 2.
Market sectors – and case studies

Market sectors – some example case studies of the use made of the 2001 Census
Government – National
• Commission for Rural Communities – Deprivation in rural areas
• Department for Transport – Accessibility Indicators for local transport planning
Government – Local
• Newport City Council – Community Development – evaluating funding and outcomes
• Suffolk County Council – Rural Enterprise Scheme
Health Service
• UCL & Department of Health – Strategic Review of Health Inequalities in England
• Yorkshire & Humberside Public Health Laboratory – Use of a range of

Market sectors – some example case studies of the use made of the 2001 Census

neighbourhood geodemographic classifications

Academic

- Newcastle University – Defining local labour markets, and other functional regions, for government policy purposes

Commercial

- NOP – Optimising survey and sample design
- Sainsbury's supermarkets – Planning of store investments and developing network strategy

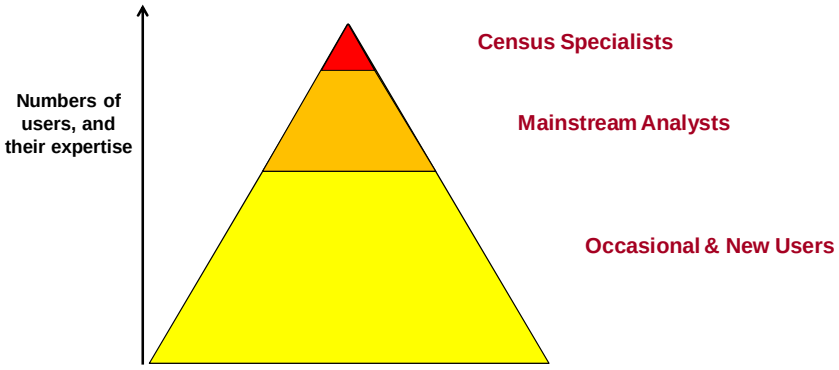
Charity

- Carers UK – Developing improved policies for combining work and care

B. Segmentation – variations in users' knowledge, and time available

13. It is also vital to recognise the great variations amongst individual users. In recent years the number of census specialists, who devote much time to developing their expertise, has increased. However, there are many more users who consider themselves general data analysts, rather than census experts, and who may spend a few hours – but no more – in exploring the census for a particular project. Thirdly, there are yet more occasional and new users, who seek simple summaries that they can access easily and swiftly.

Figure 1.
Census users – the pyramid



V. We can learn from Value-Added Resellers

14. In addition, it is necessary to recognise the importance in the United Kingdom of the census Value-Added Resellers (VARs), such as Experian and CACI. These are service companies which make it easy for clients to buy selected datasets, geodemographic classifications, and more complex analyses. They are particularly used by commercial

companies, but some central and local government organisations also find their services valuable.

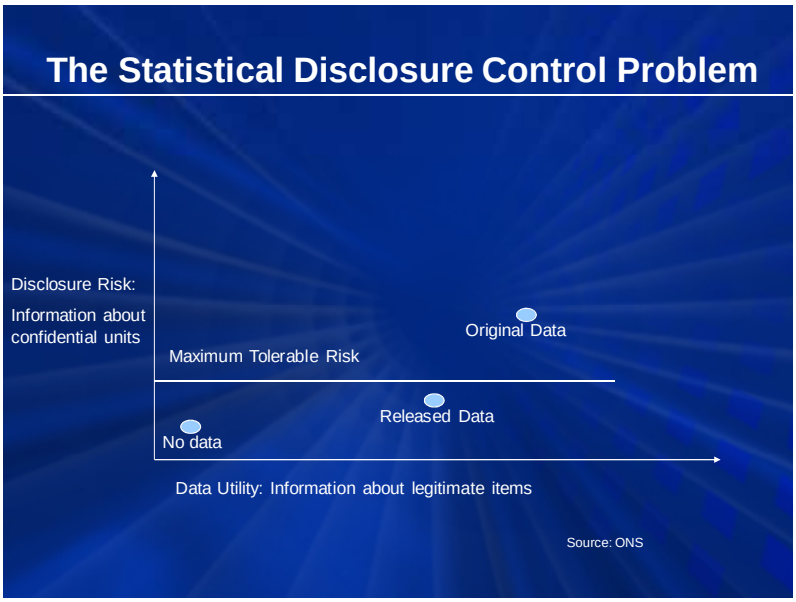
VI. Maximising use and value (1) – by seeking a better balance between risk and utility

15. How can Census Offices maximise use and value? Before considering a range of ways in which the pyramid can be grown, it is important to highlight one issue which has caused much concern amongst users in the United Kingdom: statistical disclosure control (SDC).

16. The Census White Paper[6] rightly puts a strong emphasis on the importance of statistical confidentiality, and outlines several measures to ensure disclosure control "that is, to prevent the release of statistical information that identifies characteristics about an individual person or household."

17. Since the release of the 2001 Census, the ONS has continued to put considerable research effort into minimising disclosure risk without excessive negative impact on data utility. The problem has been illustrated with the following chart:

Figure 2.
The Statistical Disclosure Problem



Source: Jane Longhust & Paul Jackson, ONS[7]

18. SDC has an extensive technical literature, and a range of possible solutions, which include "safe data" (perturbation, limited subject and geographical detail, etc.), "safe (approved) researchers"; safe settings"; and "safe licences".

19. Considering the diagram above, the ONS's research interprets utility as the amount of information detail that is made available, for example: "The utility of microdata that has

[6] http://www.official-documents.gov.uk/document/cm75/7513/7513.asp
[7] www.rss.org.uk/rssadmin/uploads/142133_ONS%20Jackson%20Longhurst.ppt

undergone Statistical Disclosure Limitation methods is based on whether statistical inference can be carried out and the same analysis and conclusions drawn on the perturbed data to the original data."[8] This approach was developed when ONS was seeking to establish the extent to which data could be damaged without rendering it useless.

20. However, there is another view of utility:

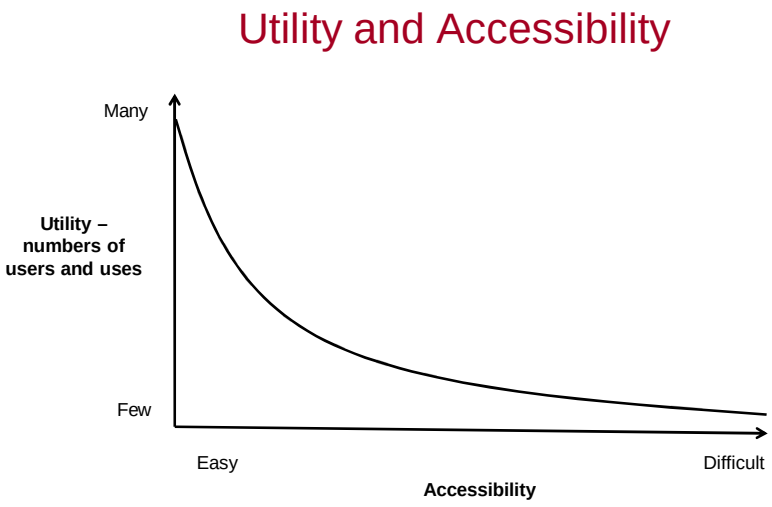
> *"That property in any object, whereby it tends to produce benefit, advantage, pleasure, good, or happiness……. It is the greatest good to the greatest number of people which is the measure of right and wrong."*
>
> Jeremy Bentham. Introduction to the Principles of Morals and Legislation (1789)

21. This interpretation seems closer to the Code of Practice and its requirement to "Ensure that arrangements for confidentiality are sufficient to protect the privacy of individual information, but not so restrictive as to limit unduly the practical utility of official statistics." The UK Census Offices are currently deciding how best to apply SDC in 2011, and it is to be hoped that both statistical and practical utility will be considered.

VII. Maximising use and value (2) – a wider manifesto

22. Widening from the particular issue of SDC, in what other ways can the Census Offices maximise the use and hence the value of all the information that is collected, and grow the pyramid of users? The following recommendations, which are based on experiences in the United Kingdom, focus on the numerous occasional and new users at the base of the pyramid, seeking Bentham's greatest good to the greatest number of people, with the emphasis on accessibility (Figure 3).

Figure 3.
Utility and Accessibility



[8] http://www.amstat.org/sections/srms/proceedings/y2008/Files/300242.pdf

A. Terms and Conditions

(a) Free at the point of use. The UK's 2001 Census achieved this for all users for the first time by central funding.

(b) Simple licensing. Again, the 2001 Census broke new ground with its very simple Click / Use license.

B. Statistics – specification:

23. As well as producing detailed cross-tabulations for administrative areas such as municipalities (as traditionally required by government) many more new users can be attracted by:

(a) Statistics for very small areas. In the United Kingdom these are Output Areas of c.120 households. Such areas are often of interest in themselves, but most importantly can be aggregated to create new ad hoc areas such as areas of deprivation, or retail catchments. Such small area statistics are of immense value to users.

(b) Simple key statistics. New and occasional users find it much easier to cope with selections of perhaps 100 key univariate statistics than complex tables: less means more.

24. Simple geodemographic area classifications, which again help to reach and create new markets.

C. Statistics – delivery

(a) The internet has transformed the potential for easy access to statistics, with many datasets downloadable, and only the largest ones needing to be supplied on DVD.

(b) It is, however, important for new users to be offered popular formats, such as Excel and csv: the UK Census Offices' initial adoption of SuperTable in 2001 did not meet many users' needs.

(c) In addition, organisations (such as the academic community, and VARs) which act as distributors of detailed statistics for the whole country, needed special bulk downloads in 2001. Plans for 2011 include the development of an Applications Programming Interface (API), which might also create new markets amongst popular search websites.

(d) Email alerts to let users know of new releases, not just putting datasets on the website.

D. Accompanying geographical infrastructure:

25. As well as statistics, the following are of great value to users:

(a) Digital boundaries for Output Areas: these became a free part of the Census package for the first time in 2001, and have been immensely useful, both for the Census and also for subsequent datasets created from new sources.

(b) Postcode/Output Area directory: again, free for the first time in 2001, enabling users to aggregate and profile administrative and customer files.

(c) Map background: this was hampered by licensing restrictions in 2001, but the recent www.data.gov.uk initiative has now put high quality mapping into the public domain.

E. Visualisation

(a) There can be no doubt that maps are immensely powerful in drawing people's attention to census statistics, and expanding the marketplace.

(b) Mash-ups, combining external data sources with Census statistics, are rapidly increasing in popularity – for example Maptube: http://www.maptube.org/

F. Comparability across counties – of the United Kingdom, and Europe

(a) Many users seek to compare variables across national boundaries, and much time and frustration could be saved if the Census Offices make it clear which variables use common definitions (and also those which cannot be compared), and produce comparable packages of key variables.

(b) Users in the United Kingdom are also hoping that such packaged datasets will all be available from a single website, rather than having to be consolidated country by country. We are looking forward to this extending not just across the United Kingdom, but to Europe as a whole.

VIII. Conclusion

26. We must keep Jeremy Bentham firmly in mind!