W1.2

2001 Census: its radical innovations and benefits for users

Keith Dugmore, Demographic decisions

1 Introduction

Use of Census data has exploded in the last two decades. Government has always used it for allocating resources to local areas, but the major growth has been in the marketing world. Market researchers use the Census when designing samples of populations. Market analysts use geodemographic classifications to segment and target their customers and prospects. They also regard the Census as a foundation of area analysis, assessing store locations and targeting local marketing campaigns. The Census underlies many decisions which involve investments of billions of pounds every year.

The 2001 Census held on 29th April will provide unprecedented detail about the United Kingdom's current population. Since the last full count ten years ago, many sample surveys have pointed to significant social changes at both national and regional levels. The new Census will provide counts and classifications of the population right down to the smallest neighbourhood.

This year's Census has several major innovations, including new questions, the use of postal geography, and new classifications, and is using automatic data capture and coding technologies. Dissemination of the results will be revolutionised. All these developments will open up new opportunities for tens of thousands of users in both public services and commercial companies.

2 What's new in 2001?

Users have been in active consultation and dialogue with the Census Offices about plans for 2001 since the mid 1990's. Public sector users in central government, health, local government, and the universities have been well represented. From the private sector, the Market Research Society's Census Geodemographics Group has had an active role in since 1991, and the Association of Census Agencies has represented suppliers such as Experian and CACI. More recently, the Demographics User Group (DUG) now represents 14 large commercial users ranging from Abbey National and Bank of Scotland to Whitbread and Yellow Pages.

The Census traditionally asks a series of core questions such as age, sex, marital status, country of birth, occupation, industry, qualifications, housing tenure, rooms, and car ownership. Address one year ago identifies migration, whilst the questioning of workplace enables the analysis of commuting flows.

This time there has been an updated question on ethnic group and new questions on religion, general health, lowest floor of accommodation, and the provision of unpaid personal care. The topic of religion illustrates that, although the Census is carried out for the United Kingdom as a whole, there are some differences in the questions asked in each country. The questioning of religion was also unusual in that it was voluntary. Income is the one vital topic that was not be asked: although users from both public and private sectors made a strong case, the government ultimately decided not to take the risk of including a question that might affect response to the census as a whole.

One of the Census Offices' key objectives has been to maximise the coverage of the whole population. This time, the aim was to count all usual residents – the population base did not include visitors, who were counted at their address of usual residence if this was in the UK. The census form was redesigned to

encourage full response. One further innovation, the posting back of Census forms, rather than having them collected door-to-door, was feared by some users as a likely cause of reduced response. Happily, this has been unfounded, and the Census Offices believe that the response has been at least as good as 1991's 98%.

The innovations in geography are fundamental. The collection areas, or enumeration districts, in England and Wales were planned using a Geographic Information System together with Ordnance Survey's ADDRESS-POINT product, which gives every postal address in the country a grid reference to 1 metre, to produce maps and address lists for enumeration. It also provides a base for aggregating the data from the forms by either postcode or grid reference.

Most significantly for users, the Output Areas for which aggregate statistics will be published will also be built from postcodes, but designed quite separately from the collection areas: we will have broken away from the traditional practice of using enumeration districts for both collection of forms and the publication of statistics. Moreover, the new Output Areas will be smaller, will respect administrative boundaries, and will be designed to be as homogeneous as possible in terms of variables such as housing tenure. Typically, Output Areas will contain 100-120 households, totalling c.200,000 across the United Kingdom as a whole. The statistics will also be accompanied by digital boundaries for mapping. In Scotland, the trail blazed in the last two Censuses of using postcodes to build enumeration districts and separate Output Areas will continue.

The other great innovation for users will be in the coding of the forms. In previous censuses, questions which have been difficult to code (such as occupation and workplace address) have only been processed for a 10% sample of the completed forms. This time the increased use of automatic scanning and data processing technologies will enable the information on every form to be fully coded, so that there will be 100% counts for every variable. The great benefit of 100% coding to users is that large sampling variability affecting statistics for small areas will be eliminated. Other novelties in coding are that Social Class and Socio-Economic Group are to be replaced by a new classification, the "National Statistics Socio - Economic Classification". More importantly for commercial users, an approximation to Social Grade will be included for the first time. There is also the probability that new household-level classifications will be introduced.

One further area of innovation is what has become known as the One Number Census. There was some concern about undercoverage in 1991, both its extent (2%) and bias (especially young males and old females). This time the largest ever post-Census Coverage Survey is being used to produce an estimate of the true total population. The Census counts will be adjusted accordingly, with missing records being imputed. A consequence of this is that statistics will be published in a series of tranches simultaneously covering areas throughout the whole country. Most users are in favour of some simple correction, but some are wary that excessive time might be spent in adjusting results which will be ageing month by month.

3 What will be the main deliverables this time?

After several rounds of consultation, we are nearing the final specification of the statistical products. Users are naturally seeking maximum detail of both geography and topics, but the Census Offices have to temper this due to the need to prevent disclosure of information about identifiable individuals.

An executive summary of the main products is:

3.1. Key Statistics. c.300 simple counts, often in the form of percentages; available for Output Areas and above, and also as printed reports at Local Authority level.

3.2. Census Area Statistics. The key dataset for most geographers, these provide detailed univariate tables and also more than 70 detailed cross-tables with c.5,000 unique counts down to Output Area level. The CAS include not only people resident in each area, but also those who work there, and those migrants used to live there 1 year ago.

3.3. Standard Tables. These comprise more detailed versions of the CAS cross-tables, running to c.100 tables with c.18,000 unique counts, but are only available for wards and larger areas.

3.4. Postcode Directory. A lookup file of Output Areas / Postcodes (some of which will be split), that will enable links to other datasets such as customer or patient records.

3.5. Boundaries. Again of great interest to geographers, Output Area digital boundaries, with grid reference centroids & area (hectarage).

3.6 & 3.7. Origin / Destination Statistics. There are two sets. Workplace statistics classify movements from home to work, and will provide at least some basic counts right down to Output Area level. The Migration statistics will classify the moves of people who lived at a different address 1 year ago; these tables go down to (only) ward level.

3.8. Sample of Anonymised Records. Taking forward a 1991 Census innovation by the academic community, this provides a file of individual anonymised records (not counts for areas). Unfortunately, the trade-off for this individual detail is that each record will only be coded to a comparatively large geographical area such as a Local Authority or perhaps ward.

In addition, users will also be able to order their own special tables. Ordering ad hoc tables from the last Census was costly for users in both money and time. For 2001 a much slicker service for ordering novel tables is being developed. Users will be able to define and submit orders over the Web. As well as specifying new combinations of standard variables such as occupation or ethnicity, there is also the prospect of defining new variables such as indices of affluence or poverty.

And when will the data start to be published? The Census Offices have not yet produced a detailed timetable, but we expect some initial results at the end of 2002, with the full Census Area Statistics in the second quarter of 2003.

4 Census Access – a revolution in dissemination

Access to the data will improve in two ways. The first is the more predictable: the huge advances in technology in the last ten years will make life much simpler for users. The Office for National Statistics is keen to distribute data over the Web, and for large national datasets we should be in the world of CD-ROM and DVD, rather than the vanload of magnetic tapes of only ten years ago.

The second revolution in access - the licensing of the data - could not have been predicted and is even more significant. For the last Census there were two approaches. Public service bodies such as local authorities and the universities each made mass purchases for all their constituent organisations; others, especially commercial companies, turned to Census Agencies, who supplied data, paying royalties to the Census Offices. In the latter case there were also restrictive licensing conditions, limiting use of the data to a fixed number of computers.

Members of the Demographics User Group made the case that access should be improved in 2001 by enabling the purchase of data directly from the Census Offices, as an alternative to Agencies. They also sought more liberal licensing arrangements, enabling use of the data freely throughout companies. Most ambitiously, the case was advanced that Census data should be made freely available, as is done in the USA.

However, the legislation which governs the Census requires the recovery from customers of the additional marginal costs of output which is not published in reports to Parliament. The Census Offices have therefore arranged a scheme called Census Access in which costs will be recovered up front from a limited number of 'funding partners', boosted by an award of new money to improve dissemination services and make them much more user friendly. The exciting result is that standard Census output (Items 3.1-3.5 above) should be accessible free in effect to all end users.

The Government is now intending to make information which it has collected for its own use - including statistical information and the Census in particular - available via the Internet for re-use at little or no cost to encourage the growth of the knowledge economy. An announcement of a new centralised "Click-Use" licensing scheme to cover re-use was made on HMSO's website in April 2001.

Where does this leave the Census Agencies? In 1991, the overwhelming majority of commercial users turned to the Agencies for their Census data. Commercial users will have the alternatives of obtaining 2001 output directly from the Census Offices as part of a new the arrangements, or of buying value added services from commercial suppliers. The Agencies, rather than being simple resellers of Census data, will now

concentrate on adding value. This appeals to most users, who don't live and breathe Census data every day, and need support, and also to companies of all types who are seeking derived data such as geodemographic classifications, additional survey and lifestyle datasets, and sophisticated analyses.

5 Neighbourhood Statistics – the bigger picture

During the last two years the Cabinet Office's Social Exclusion Unit has been promoting its National Strategy for Neighbourhood Renewal. One of its Policy Action Teams (PAT 18) produced a report "Better Information" which identified major topic domains such as Health, Crime, and Housing, and sought to assess what data is, or might be made, available for small areas. It recommended that data from administrative files, surveys and modelled sources should be made available freely on the Web, with the ONS acting as One Stop Shop, rather than users having to trawl around many separate government Departments.

ONS's new Neighbourhood Statistics service was launched on the web in February. Its start has been pragmatic, providing available data for current wards, but it is to be extended to encompass new sources, including the 2001 Census statistics, and also other data such as modelled estimates of income. Ideally these will be produced for Output Areas, enabling current administrative statistics for topics such as benefits, health and crime to be related to Census counts of population for small areas.

6 The leading-edge users: adding value

Given this avalanche of new data, what are the advanced users' plans? Large retailers and financial services companies have been planning for the arrival of the Census for the past three years. Their prime interests are in site location, and segmenting customer and prospect files. The availability of new 100% counts for very small areas, which can be related directly to postcoded files, is an exciting development. The new question on religion is of some interest, but getting good data about workplace populations is a particularly hot issue, with increasing amounts of expenditure being made near work rather than home. The arrival of social grade, and a new household classification will be widely welcomed.

Typically such companies will obtain large datasets of many variables for every Output Area in the country, together with a postcode / Output Area directory and also digital boundaries for mapping. These Census datasets will then be integrated into systems which already contain much customer data and other external datasets, such as Neighbourhood Statistics.

Advanced users now have huge warehouses of Census and other small area data, but the fundamental issue is "what information is needed to answer specific business questions?". The challenge is to add value by selecting and combining datasets to create new information. Decision-makers seek simplification
.....*"Less means more"*

Further information

The latest news of the Census from the Office for National Statistics:
www.statistics.gov.uk/census2001

Details of the new "Click-Use" licensing scheme which provides free access to much government data:
www.hmso.gov.uk/copyhome

The Social Exclusion Unit's Policy Action Team 18 report "Better Information":
www.cabinet-office.gov.uk/seu/2001/PATaudit

Neighbourhood Statistics for small areas downloadable from the web:
www.statistics.gov.uk/neighbourhood

Geodemographics Knowledge Base, providing a portal to services for targeting local markets:
www.geodemographics.org.uk

Market Research Society, which is planning to produce a Census website and to publish a Census Guide to coincide with the delivery of the first statistics: www.mrs.org.uk