

A07.3

A user's view: Why the 2001 Census will change everything

Keith Dugmore, Director, Demographic Decisions

1 Introduction

The 2001 Census results to be released in the coming months 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.

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 large commercial users.

The latest Census has several major innovations, including new questions, new classifications, and the use of postal geography. Dissemination of the results will be revolutionised. All these developments will open up new opportunities for users in thousands of organisations, both public and private.

2 What makes the 2001 Census special?

Questions

Censuses traditionally ask 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.

The 2001 Census also included 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 question on 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 is also different in that it was voluntary. Income is the one vital topic that was not 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.

Coverage / One Number

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 does not include visitors, who were counted at their address of usual residence if this was in the UK. The census form was also redesigned to encourage full response. One other innovation, the posting back of Census forms, rather than having them collected door-to-door, was feared by many users as a potential cause of reduced response, but at the end of the day the overall response rate reached 98%, similar to that in 1991.

Closely related to the issue of coverage 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 has been used to produce an estimate of the true total population. The Census counts have been adjusted accordingly, with missing records being imputed. A consequence of this is that statistics are being published in a series of tranches simultaneously covering areas throughout the whole country, rather than county-by-county. Most users are in favour of some simple correction, but some have been wary that excessive time might be spent in adjusting results which will be ageing month by month.

Coding

One of the most significant innovations for users is 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 has enabled the information on every form to be fully coded, so that there will be 100% counts for every variable; the only exception is that, for economy reasons, this will not extend to coding the previous occupation of all people who are not now economically active. 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 being replaced by a new classification, the "National Statistics Socio - Economic Classification", and, after considerable development work by the Market Research Society, an approximation to Social Grade will be included for the first time.

Geography

The innovations in geography are fundamental. The collection areas, or enumeration districts, in England and Wales have been 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 postcodes 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. 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.

Outputs . . . and Disclosure Control

After several rounds of consultation with users, the statistical products have now been specified. 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:

- **Key Statistics.** c.500 simple counts, often in the form of percentages, available for Output Areas and above, and also as printed reports at Local Authority level.
- **Census Area Statistics.** The key dataset for most geographers, these provide detailed univariate tables and many cross-tables, with more than 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.
- **Standard Tables.** These comprise more detailed versions of the CAS cross-tables, but are only available for wards and larger areas.

- **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.
- **Boundaries & Map Background data.** Again of great interest to geographers, Output Area digital boundaries, with grid reference centroids & area (hectarage), and also background mapping.
- **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 ward level.
- **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.

The one major constraint on the provision of detailed outputs is what has become known as Disclosure Control. Both the Census Offices and users accept that it should not be possible to identify named individuals by poring over the statistical tables. Until late last year the agreed approach by Census Offices across the UK was to make some modifications to the database before processing tables. The ONS then shocked users by announcing that disclosure control would be implemented in England & Wales by rounding the numbers in each table to multiples of three, with the consequence that rows and columns would not add up. After sustained pressure by users, a compromise method, applying some rounding but providing consistency within (but not between) tables, has been announced. The Census Office in Scotland has retained the original proposals for pre-tabulation modification, which will provide completely consistent output.

And when will the data start to become available? The Census Offices for England & Wales, Scotland, and Northern Ireland are acting independently, but their targets are to publish the Key Statistics in about three months' time, with the more detailed outputs flowing throughout 2003.

Access & the Census Agencies

Access to the data will improve in two ways. The first is 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.

Four years ago the Demographics User Group, which represents large commercial users, started to make the case that access should be improved by enabling the purchase of data directly from the Census Offices, as an alternative to supply by Census Agencies. It also sought more liberal licensing arrangements, allowing use of the data freely throughout companies.

In the event, the Census Offices have gone much further. The Government has now decided to make information which it has collected for its own use - including statistical information and the Census in

particular - available largely 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 licensing scheme to cover re-use was made in April 2001. ONS has a programme to take National Statistics into the 'information age' via Web based dissemination, and 2001 Census output will be accessible on this radically new basis.

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 should be accessible free in effect to all end users in line with the wider Government policy.

Since the first value-added reseller of Census output set up in 1977 as a Census Agency their numbers and activities have grown considerably. In the early days their role was simply to package Census statistics, supplying selections of variables for areas defined by users, such as radii around retail outlets, whilst paying royalties to the Census Offices. Census-based geodemographic classifications followed, and now companies such as CACI, Claritas, Equifax, Eurodirect and Experian offer many sources of data, software and analysis projects. In 1991, the overwhelming majority of commercial users turned to the Agencies for their Census data. With the free availability of the 2001 Census statistics, the agencies will clearly have to be seen to add value, rather than simply resell basic selections of data.

3 Why is the Census vital to both public services and commercial companies?

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 often 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.

4 Who will be using it?

The users of the Census are a very diverse group. One chasm that is rarely crossed is that between users in the public services and those in commercial companies. Hitherto they have existed in largely separate universes, although the move to postal geography and the free distribution of data should provide a big impetus towards the sharing of data and analytical techniques.

Users also vary widely in their sophistication, from those specialists who live and breathe Census data every day to others who grab statistics on an occasional basis. Looking back over the last 30 years, the pyramid of users has grown considerably, and the advent of free data will attract huge numbers of people who have never before used the Census.

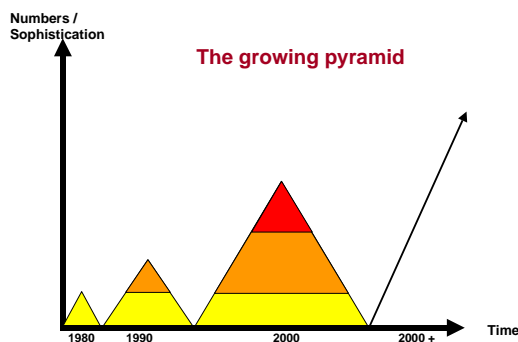


CHART 1

5 What will the advanced commercial users be doing?

Large retailers and financial services companies represented by the Demographics User Group have been planning for the arrival of the Census for the past four years. The Group's membership comprises Abbey, Boots, HBOS, Marks & Spencer, Nationwide, Royal Bank of Scotland, Saga, Sainsbury, Tesco, T Mobile, Whitbread and Yell. Their businesses typically cover the whole of the United Kingdom, and they have developed teams of specialists who are continuously hunting and analysing a wide variety of datasets.

Such companies always face some fundamental questions. Who are our best customers? Which are our best prospects? Where are the best places for our new outlets? What should we be offering in each of our existing outlets? The answers are produced by the relentless classification of customers, estimating of local markets, and evaluation of sites.

At first sight this might appear to be very different from the world of public services, but there are in fact many parallels. The classification of customers can equally be applied to library users or sports centre customers. The analysis of local markets can be applied to the demand for Day Centres or to estimating the amount and type of household refuse generated in an area. The siting of services applies just as much to hospital or council One Stop Shops as to bank branches or pizza restaurants.

Typically the large commercial users 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. The availability of 100% counts for very small areas, which can be related directly to postcoded customer 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 will be widely welcomed. Census access will enable data to be made available to all users within a company, ideally using intranets. These Census datasets will then be integrated into systems which already contain much customer data and other external datasets. The range of datasets which are now being integrated to create new information is illustrated below.

Other datasets which are being integrated with the Census by large commercial users

Customer Files (Postcoded)			
Small Area Statistics			
Census - Residence			
Census - Workplace			
JUVOS Claimants			
Income estimates			
Market estimates			
Updates			
Projections			
Digital Boundaries			
Postal			
Administrative			
TV Areas			
Map Data			
Background - OS, AA			
Road network - drivetimes			
Points & Sites			
Retail Locations			
Goald Plans			
Business Locations			
Planning Applications			

Postcode Directories & Classifications			
Codepoint			
EDs			
Acorn			
Mosaic			
Cameo			
Prizm			
Flood Risk			
Lists			
Lifestyles			
Electoral Roll			
Household classifications			
Person classifications			
Suppression files			
CCJs			
Businesses - CoHo & Yell			
Sample Surveys			
FES			
Target Group Index			
AGB SuperPanel			
Shopping catchments			
Traffic flows			
Pedestrian flows			

CHART 2

In addition, the new Neighbourhood Statistics service will provide modelled estimates of income, and administrative statistics for topics such as benefits, health and crime, which can also be related to Census counts of population for small areas.

6 The Census will change everything

Why will the 2001 Census change everything? The data will be free, as part of the Neighbourhood Statistics service. The new Output Areas will provide a common geography based on postcodes, with boundaries and directories to link other datasets. Consequently, permutations of data to create new information will be infinite. Use of the Census will spread and deepen within established user organisations, whilst new users – small organisations and individuals – will be attracted in large numbers.

Further information

The latest news of the Census from the Office for National Statistics:

www.statistics.gov.uk/census2001/default.asp

The Census in Scotland:

<http://www.gro-scotland.gov.uk/grosweb/grosweb.nsf/pages/censushm>

Details of the new “Click-Use” licensing scheme which provides free access to much government data:

www.hmso.gov.uk/click-use-home.htm

Neighbourhood Statistics for small areas downloadable from the web:

www.neighbourhood.statistics.gov.uk/home.asp

Geodemographics Knowledge Base, providing a portal to services for targeting local markets:

www.geodemographics.org.uk/