

ODUG response to the Office for National Statistics (ONS) consultation on the Census and future provision of population statistics

Introduction

The Open Data User Group (ODUG) welcomes this opportunity to respond to the Office for National Statistics (ONS) consultation on the Census and future provision of population statistics in England and Wales.

Carried out every 10 years, the Census is the primary source of population data at national and local level in the UK, the closest dataset to a complete 'population register'. The Census is rightly recognised as part of the critical National Information Infrastructure (NII) ¹.

Following the Policy Action Team review on Better Information and Crown Copyright Review of 1998, Census data (statistical tables rather than raw data), which had since the Rayner Review of 1982 been treated as Tradable Information and locked behind a pay-wall, was released as open data along with other statistical information from ONS such as Neighbourhood Statistics. The use, value, and the range of uses of Census Data, have been greatly increased as a result of it being freely available. Because of this the damage that would be caused by the loss of the Census, or a workable open alternative, would be substantial.

Changes to the Census are therefore undertaken with great care. The current options under review potentially create the greatest change in how the Census is undertaken in the last 150 years.

The options for Census taking in future

The consultation asks users to respond on the two primary options for future Censuses identified by the ONS 'Beyond 2011' programme:

Option 1. Online Census every decade. This would essentially retain the features of the current system, with a complete Census of the population carried out once per decade. Data outputs would also remain the same, with outputs for the majority of population statistics published down to Output Area² once per decade. Costs for this option are estimated by ONS at £625 million over the decade (£1.10 per person per year).

Option 2. Annual outputs based on government administrative data sources plus an annual 'rolling' Census of 4% of the population. This would produce more regular population statistics, with annual updates to data for Local Authorities and higher areas. Additional data outputs could also be

¹ Cabinet Office (2013). National Information Infrastructure, <https://www.gov.uk/government/publications/national-information-infrastructure>

² Output Areas are the smallest building block used by government for statistical open data. There are roughly 180,000 Output Areas across England, with an average population of 300 people, or 120 households.

provided from the administrative datasets used, including measures of local incomes (the primary socio-economic dataset missing from the current Census). However the most detailed local area statistics would be lost under this option, with only population by age and gender published to OA level. Costs for this option are estimated by ONS as £460 million over the decade (80 pence per person per year).

Benefits and value of the Census

The publication of the 2001 Census data as freely available and reusable by all users is a major milestone in the development of open data across the UK, well before the term 'open data' came into use. Costs to access the data from previous Censuses were expensive, with licensing terms restricting onward publication and reuse. Publishing the 2001 Census as open data led to very widespread use of local area statistics, such that it would now be difficult to find a public organisation or service provider that did not rely heavily on Census data, or data based on Census, to target resources to local areas.

Key uses for the Census broadly fall into three broad categories:

1. Population data for targeting resources & services. The data from Census delivers a set of population statistics that are used directly by public, commercial and academic sectors. For example: government funding allocations; location of new services such as hospitals, schools, supermarkets; (re)definition of voting boundaries such as parliamentary constituencies.
2. Population denominators and inputs for calculation of other statistics. Census data is also a component in the statistical models used to create a range of public and private sector data outputs, including the great majority of National Information Infrastructure datasets in the 'Social Mobility & welfare' and 'Society & culture' themes³. Other examples include: calculation of annual mid-year population estimates, population denominators and socio-economic indicators for the Index of Multiple Deprivation; inputs to commercial data outputs such as Acorn, Mosaic; Weighting national and local surveys to reflect population characteristics.
3. Analysis of demographic, social & economic trends. The complete and detailed coverage of Census data enables robust analysis of social issues. For example, the early results from Census 2011 confirmed dramatic population increases over the last 10 years, as well as shifts in the social and demographic make-up of local areas.

It is difficult to place a rigorous economic value on high quality population data from the Census, although this value is undoubtedly high. However it may be more appropriate to consider the resources allocated using Census data and subsequent mid-year estimates. The Revenue Support Grant allocation to Local Government is almost £24 billion, and makes widespread use of Census data directly, as well as indirectly through population datasets modelled on the Census (such as projected residential populations)⁴. Other datasets that make use of Census indicators are used to allocate specific funds; for example University of Oxford has estimated that the Index of Multiple Deprivation is used in the allocation of roughly 1% of total government funding.

³ Cabinet Office (2013). National Information Infrastructure datasets, available from <https://www.gov.uk/government/publications/national-information-infrastructure>

⁴ Department of Communities and Local Government (2013). Calculation of 2013-14 Formula Funding. Available from <http://www.local.communities.gov.uk/finance/1314/settle.htm>

The correct allocation of these funds, and the identification of the small areas of maximum deprivation where the need for funding is greatest, both depend on accurate local statistics.

Future provision of population statistics should include more timely and robust data outputs

The primary drawback with the 10-yearly Census option is the lack of timely open data on population changes. Provision of more regular data is a key advantage of the administrative and rolling Census data option.

It is remarkable that local government funding settlements agreed for 2013/ 14 – distributing some £24 billion - are still using 2001 Census data indicators. Much has changed in the 12 years since this data was current.

Similarly, not until 2011 Census data was published in 2013 did we have data on the full impact of migration from the EU Accession countries from 2004 onwards. Of course, datasets such as National Insurance registrations from overseas, birth registers, school Census and so on each gave a partial picture of the changes, but only Census gives the full picture across the country at local level. Again, timely data is needed to enable services and resources to respond to changes in local populations and needs.

In the 21st century, we would expect to see regular publication of the key population statistics used to allocate large proportions of government funds, and target public and commercial services. It seems archaic to wait for the 10 year survey to obtain population statistics that are seen as reliable enough for allocating government funds. ODUG believes that future provision of population statistics should include more timely and robust data outputs.

Small area data is a critical – and valuable – population statistics output

The administrative data option will provide more limited small area data, which is a major drawback to this option⁵. This is particularly important as “troubled families” which cause the greatest cost to the public purse, and have the greatest impact on their communities, are known to be concentrated in small geographical areas. While many of these remain constant over time, there is a level of geographical change in the most deprived areas that has been tracked by the IMD and requires more detailed local statistics than an administrative alternative to the Census can provide.

Quite apart from the most deprived areas, public and commercial organisations increasingly need fine-grained data to target services to the right areas and groups. Decisions on locating services such as hospitals, supermarkets and schools, need to be underpinned by more detailed information than local authority level data. Similarly, investment in to local areas by groups such as Big Lottery (£600 million per year⁶) and social housing associations (£700M per year ‘community investment’⁷), as well as new models for local financing such as Social impact bonds⁸, require detailed small area statistics to identify areas most in ‘need’.

⁵ “A limited set of basic estimates by age and sex would be available for small groups of postcodes, but the most detailed statistics for Output Areas that the Census has provided in the past would not be available using this method”, ONS (2013), Census consultation document.

⁶ <http://www.biglotteryfund.org.uk/about-big/our-approach/about-big-lottery-fund>

⁷

⁸ See <https://www.gov.uk/social-impact-bonds> for an overview of Social impact bonds.

In the great majority of these cases, Census data provides the small area population counts and denominators used in the mechanisms for allocating resources. Data is typically required at the most detailed Output Area (OA) or Lower Super Output Area (LSOA) level for these calculations, in order to construct estimates for specific geographic areas such as 'catchment areas' or 'service patches'; Local Authority data is not sufficient.

The Index of Multiple Deprivation (IMD) provides a ready example of the importance of small area Census data. The IMD is a measure of deprivation based around seven domains of deprivation – low income, worklessness, education, health, crime, environment and barriers to housing and services each of which is based on a set of indicators from a range of government administrative datasets, surveys and models. Published at LSOA level across England, with similar measures across Wales, Scotland and Northern Ireland, the IMD makes heavy use of LSOA level Census data for population denominators, indicator modelling, as well as direct use of Census indicators of adult skills and housing condition. Oxford University has estimated that 1% of all government funds were allocated at least in part on the basis of the IMD⁹, and it is difficult to find a local authority or agency that does not make use of the index in some form.

Given the current emphasis on the value and power of data for public services, economic growth and transparency, it seems both short-sighted and unambitious that data availability for neighbourhoods across the UK would be less good in 2025 than in 2005. It would be a major step backwards to downgrade the OA and LSOA level data in future population datasets - as suggested in Option 2 (administrative and rolling Census survey).

There are benefits to the administrative data & rolling Census option, but further real world testing is needed to confirm that this option does not unacceptably reduce the quality of available data

The risks and benefits for Option 1 (decennial Census, with the 2021 Census primarily online) are well understood. This is essentially the 'as you are' scenario, plus cost savings and quality improvements through the move to online.

The primary benefits for Option 2 (annual outputs, based on administrative data and rolling 4% Census) are threefold: 1) more regular updates for the primary data outputs; 2) opportunity to include measures such as income that are not captured by the Census at present; and 3) reduction in costs over the decade.

Option 2 therefore suggests potential for faster, better and cheaper – a step change in the provision of population statistics. However, ODUG agree with the Royal Statistical Society that significant unknowns and risks remain with this Option¹⁰. The case has not yet been made to be clear that the outputs from this option would be sufficiently robust to replace the current complete Census option.

Without the complete Census to act as a cross-check validation, a clear risk is that errors in annual estimates and methodologies could remain unidentified. As an example, the 2011 Census data publication highlighted errors in the annual mid-year estimate series, with some local authorities (notably in London) seeing increases of 30% in their population estimate from 2010 (based on the mid-year estimates) and the 2011 Census. This has more than simply academic interest; many local authorities have seen their year-on-year funding allocations significantly underweighted.

⁹

¹⁰ Royal Statistical Society (2013). Statement on the future of the Census in England and Wales. Available from <https://www.rss.org.uk/site/cms/newsarticle.asp?chapter=15&nid=127>.

It is a particular concern that operational changes in the administrative processes generating the data which will substitute for the Census may become necessary and will outweigh the need for a consistent statistical series. The proposed annual sample Census could track the effect of such changes, however.

The lack of a robust up-to date and open national address register is a critical gap in the infrastructure needed for a rolling annual Census. Echoing ODUG's recent submission to the Public Administration Select Committee inquiry into Statistics and Open Data, this re-emphasises the need for an Open National Address Dataset to underpin public services and economic growth.

Census data should also include adjustments based on administrative statistics to improve data quality

Administrative data has a part to play in acting as a cross-check to Census data. For example, 2011 Census data on social housing occupation showed that up to half a million people in the Census misreported their housing as council owned rather than registered social landlord¹¹.

Dual approach combining the best elements from the two options

ODUG believes the strongest option – that maximises value of open data published through the Census in terms of robust & reliable, timely, and local (small area) population statistics – is to maintain Option 1 for Census 2021. And to implement parts of Option 2 that enable further testing and development of the administrative data. This dual running approach is discussed only briefly in the consultation paper, and will inevitably increase costs.

¹¹ In 2011, Census data identified 1.8m properties run by registered social landlords in England, well below the 2.3m identified in housing provider returns to the Tenant Services Authority. Similarly, the 2.1m council owned properties from 2011 data is well above the 1.7m identified in Department of Communities and Local Government data.