

# Evaluating the potential for moving away from a traditional census

Becky Tinsley<sup>1</sup>, Louise O'Leary<sup>2</sup>

<sup>1</sup> Office for National Statistics, Southampton, UK; [becky.tinsley@ons.gov.uk](mailto:becky.tinsley@ons.gov.uk)

<sup>2</sup> Office for National Statistics, Southampton, UK; [louise.o'leary@ons.gov.uk](mailto:louise.o'leary@ons.gov.uk)

## 1. Background

### *1.1 Outcomes from the Beyond 2011 Programme*

In May 2010, ONS began the 'Beyond 2011 Programme' to review the future provision of population statistics in England and Wales in order to inform government and Parliament about options for the next census. In particular the programme focused on the potential to replace the census with statistics based on administrative data<sup>1</sup> already held by government, supplemented by ongoing household surveys.

Significant progress was made in acquiring administrative data and developing new methods. During that time, the Programme acquired a number of administrative data sources, which were used to produce administrative data based population estimates by Local Authority for 2011, and compared with the 2011 Census results as a measure of assessing the quality of the administrative data.

ONS carried out a [public consultation](#) in 2013 seeking views between the two front-running options for 2021 – an online census or an alternative based on administrative data plus annual surveys. The Government response to the consultation emphasised the value of the

---

<sup>1</sup> Information gathered by government departments and agencies during their day-to-day activities when delivering a service, registering people or carrying out transactions.

information provided by the census. An [independent methodological review](#) of the options concluded “*the online census option provides a methodologically sound basis for replacing Census 2011 methodology. We do not believe that there is yet sufficient evidence to support ....the administrative data option.*”

On the basis of the research and evidence collected, the then National Statistician [recommended](#), in March 2014:

- “*An online census of all households and communal establishments in England and Wales in 2021 ... [with] ... special care taken to support those who are unable to complete the census online; and*
- *Increased use of administrative data and surveys in order to enhance the statistics from the 2021 Census and improve statistics between censuses.*

*This would make the best use of all available data to provide the population statistics which England and Wales require and offer a springboard to the greater use of administrative data and annual surveys in the future. ...[It] may offer a future Government and Parliament the possibility of moving further away from the traditional decennial census to annual population statistics provided by the use of administrative data and annual surveys.*”

This approach was endorsed by Francis Maude, Minister for the Cabinet Office, in the Government’s [formal response](#) to the recommendation in July 2014 which highlighted the Government’s ambition “*that censuses after 2021 will be conducted using other sources of data and providing more timely statistical information....[subject to] sufficiently validating the perceived feasibility of that approach.*”

## **1.2 What information does a census provide?**

The census collects information on housing (such as number of rooms; type of central heating) and provides three types of information on the population:

1. the size of the population – how many people in each area, by age and sex

2. households and families – such as numbers, size and type of families; for example single-parent families
3. population characteristics – such as information on ethnicity, educational attainment, hours of unpaid caring provided, religion, etc

The census is unique in providing information on the whole population at a point in time, at small areas (for example down to areas of 125 households), and for small population groups (for example under 18s doing more than 50 hours of unpaid caring a week). It is this unique information which makes the census so valuable to government, local authorities, businesses and other users, and why there is a continuing need for the information provided by the census.

### ***1.3 Ambition to meet user needs***

ONS is aiming to replicate the information collected through the census with administrative data already held by government, supplemented by surveys. The goal is to be able to compare outputs based on administrative data and targeted surveys against the 2021 Census to demonstrate to government and other users that the alternative can produce high quality information at a lower cost, and can do so on a more regular basis.

Given the importance of producing accurate statistics, it would have been high risk to move straight to such a system without benchmarking new methods against the 2021 Census. This is in line with practice in other countries who have made the move more gradually.

This work addresses the Government ambition, described in Section 2.1. It is also in line with [ONS's Strategy](#) to be at the forefront of integrating and exploiting data from multiple sources, making greater use of administrative data across all statistics. This stance was supported by the recently published [Independent Review of UK Economic Statistics](#) which recommended that ONS “make the most of existing and new data sources...” (referring both to administrative and private sector data) to improve economic statistics.

Research to date suggests it won't be possible to produce the full range of Census statistics from just administrative data alone – it will need a mix of data including administrative data and data from surveys, and also 'big data' (such as mobile phone data) and data held by the private sector (such as private sector data on housing).

Some data collected in the census are not available in administrative sources (for example hours of unpaid caring) and surveys alone will not provide the detail that users need to measure change over time (such as being able to spot changes over a decade in unemployment rates by ethnicity for small areas). To overcome these challenges, and to produce outputs covering a similar range of population characteristics to those that a census can deliver, ONS will need access to more administrative data, and this will need to be integrated with annual surveys. This approach is described as an Administrative Data Census in Section 2.

The major constraints on how far and fast this will progress are the availability of good quality administrative data covering the range of census topics and the ease of access to these sources for ONS. We are driving forward access to further datasets using existing legislation but this requires approval by both Houses of Parliament on a case by case basis and is slow to achieve. The UK Government recently consulted on a range of proposals for new legislation to enable [\*Better Use of Data in Government\*](#), including for ONS's role in producing National and Official Statistics and statistical research. Such new legislation would significantly improve the current arrangements, enabling ONS to have a presumed access to administrative data held by government for statistical purposes, as well as enabling access to private sector data for statistical purposes. The consultation closed on 22 April 2016.

## **2.0 What is an Administrative Data Census?**

It is ONS's ambition to produce the type of information that is collected by a ten-yearly census (on housing, households and people) from an Administrative Data Census. Doing this will require a combination of:

- record-level administrative data held by Government

- a population coverage survey
- a population characteristics survey

### ***2.1 Record-level administrative data held by Government***

ONS will need access to a range of data held by Government. Access will be required at a record-level to enable these sources to be linked together. High quality linking requires name, address, data of birth and sex (as described in more detail in [Matching Anonymous Data](#)) as combinations of these variables can be used to produce links that are made with a high-level of certainty. Linking together multiple sources will improve the quality and coverage of the outputs that can be produced, and will support the production of cross-tabulated outputs, for example employment by qualifications at small geographic levels within a local authority.

### ***2.2 Population Coverage Survey (PCS)***

Previous research in the Beyond 2011 Programme identified the need to conduct a coverage survey, similar to a Census Coverage Survey, to measure and adjust for under- or over-coverage on administrative data and to enable the production of high-quality statistics about the size of the population. A PCS may cover approximately 350,000 households on an annual basis, as described in [Paper M8](#). Further work is required to refine the detail of the survey and the methods to subsequently produce estimates using the PCS and administrative data.

### ***2.3 Survey to produce estimates about characteristics of the population and households***

Research in Beyond 2011 focused on producing statistics about population characteristics using a survey that would cover approximately 4 per cent of the population (around 900,000 households) each year. In the [public consultation](#) that was carried out in 2013, users told us that this type of survey alone would not provide the granularity of information that they need to measure change over time (such as being able to spot changes over a decade in unemployment rates by ethnicity for small areas within an LA).

While surveys alone might not be able to produce direct estimates for the cross-tabulated outputs at small area levels that users require (as outlined in [Paper M12](#)), administrative data can be used in model-based approaches to improve the precision of survey estimates particularly where the survey sample size is too small for direct estimates, for example certain minority ethnic groups. There are different approaches to doing this, but a common feature of these methods is that they use relationships between the data and the target characteristic to produce estimates. This could be a direct relationship (for example, using ethnicity information from both administrative and survey sources) or an indirect one (for example, using ethnicity information from administrative sources to produce estimates on main language or religion from the survey as there may be a relationship between these variables and ethnicity).

For characteristics that are available on administrative data, the survey would need to measure and adjust for under- or over-coverage on administrative data, in a similar way to the PCS.

A survey could also be used to provide direct survey estimates for topics that are not available on the administrative data (such as hours of unpaid caring). For such topics, it might be possible to produce estimates only at LA level.

The precise design (including size) of this survey will depend on ONS's access to administrative data and an understanding of its statistical quality.

#### ***2.4 Other sources of data***

Other data (such as 'big data' or private sector data) may also be necessary to produce some types of information traditionally produced by the census, such as commuting flows by using data from mobile phones, or information on tenure by using information from property websites.

These different sources of data will need to be linked together and used in combination with a range of methods and modelling techniques in order to produce the type of outputs that users

require. This approach may additionally offer opportunities to provide users with the outputs that they need on a more frequent basis than provided by a ten-yearly census.

### **3.0 What needs to be in place to move to an Administrative Data Census?**

There are four key challenges to delivering an Administrative Data Census:

1. Accessing the range of data needed to produce outputs that are currently provided by the ten-yearly census
2. Linking together lots of independently collected data accurately whilst preserving the privacy and security of the data
3. Developing methods that can transform the linked data into outputs that meet the needs of users
4. Making an Administrative Data Census acceptable to key stakeholders, for example by providing value for money, and providing reassurance that data will be kept safe through this approach

To address these challenges, the following would need to be in place.

#### ***3.1 Rapid access to existing and new data sources***

To maximise the breadth and quality of statistics that could be provided by an Administrative Data Census, ONS would need to have rapid access to new and existing data sources from across Government. This would also need to extend to other sources of existing data that would add value. ONS would also need to be consulted before changes are made to the administrative data that may affect the quality and stability of outputs from an Administrative Data Census over time. New legislation, such as that described in the recent public consultation on [Better Use of Data in Government](#), would offer a solution to these requirements.

#### ***3.2 The ability to link data efficiently and accurately***

All countries that have moved away from conducting a five- or ten-yearly census have moved to a register-based census that is underpinned by a population register and usually an ID card scheme. This usually means that administrative data can be linked to the register(s) through a unique ID number, resulting in highly accurate linking. These registers also aim to provide

complete coverage of the population, which administrative data does not provide.

In the UK we do not have ID cards or a population register. Instead, as described in Section 2, an Administrative Data Census would involve linking together multiple administrative data sources and surveys to produce statistics on the range of topics that the census currently produces. This is not a simple task.

Individuals in the UK do not have a single unique reference number that is carried across all government-held data, making this linkage challenging. For example, data about tax and benefits from DWP and HMRC use the National Insurance Number, while GP Register data uses NHS number and School Census uses a unique pupil reference number. We therefore need methods that can link together these independent data sources accurately to enable the production of high quality statistics. An additional challenge is to do this while preserving the privacy and security of the data.

### ***3.3 Methods to produce statistical outputs of sufficient quality that meet priority information needs of users***

Accessing and linking data is only part of the puzzle. We need to deliver methods that can transform the linked administrative and survey data into statistical outputs that meet priority information needs of users. This means providing statistics on the topics that users need, at the right level of detail (for example, for small areas), and at the right quality. In response to a [public consultation](#) in 2013, users told us that we need to develop statistical methodologies that:

- provide robust estimates about the size of the population and the number of households
- provide estimates about population characteristics at a point in time to allow similar areas to be compared with one another
- provide the granularity of information that users need to measure change over time (for example being able to spot changes over a decade in unemployment rates by ethnicity for small areas)



Another key area, as outlined in Section 2, is developing the detail of the surveys that will be required and the methods to model from surveys and administrative data.

### ***3.4 Acceptability to stakeholders (users, suppliers, public and Parliament)***

In order to successfully move to an Administrative Data Census in the next decade, users of the data, data suppliers, the public and Parliament need to be convinced that this approach meets their needs. Acceptability to these four key stakeholders will be influenced by ensuring that:

- key information needs of users are met
- data is held, processed and linked while providing privacy, confidentiality and security safeguards

### ***3.5 Value for money***

An Administrative Data Census will need to demonstrate that it provides value for money compared to a ten-yearly census. This means showing either that it can deliver the benefits that users get from a ten yearly Census at a lower cost, or that the cost saving is sufficient to justify lower benefit. For example the Administrative Data Census may not be able to deliver all the outputs that a ten-yearly census provides but it may include additional benefits such as more timely, frequent data and new outputs that are not currently provided by a ten-yearly census. This is the key trade off that will need to be taken into account.

### ***3.6 How will ONS know if an Administrative Data Census is possible?***

For the government to make a decision after 2021 about the future of the census, ONS needs to provide evidence to show whether or not an Administrative Data Census is a viable approach to census-taking. In order to do that, ONS plans to do the following:

- Make progress in acquiring new administrative data sources, prioritising data sources that relate to, or may provide insight on, key topics that are currently produced by a ten-yearly census. For new data sources, record-level comparisons can be made with

the 2011 Census, which provides a good benchmark of the statistical quality of the administrative data. For example, it can highlight whether an administrative source has coverage issues, or lags in updating address information. Comparisons with other data sources can also be useful to understand statistical quality. ONS will publish an update on the progress in acquiring data each year.

- Publish Administrative Data Research Outputs on an annual basis. Annual research outputs will demonstrate the type and quality of outputs that could be produced from an Administrative Data Census. The first set of [Research Outputs](#) were published in October 2015. The range of topics will be expanded in future releases, depending on the availability of data and its statistical quality.

A key aim of these outputs is to allow users the opportunity to provide feedback on the data and on the methods used to help focus future developments. A [short summary](#) [<<insert link when available>>](#) of the feedback received from the last set of outputs is published alongside this paper.

- Conduct an annual assessment of ONS's ability to move to an Administrative Data Census. This will ultimately conclude with a comparison of combined administrative data and survey based outputs against the 2021 Census outputs to benchmark this approach. This will culminate in a recommendation in 2023 on ONS's ability to switch to an Administrative Data Census.
- Have methods and research reviewed by an external expert panel. These reviews are currently planned to take place in 2017, 2020 and 2022.

In May 2016, ONS published the first assessment of its progress towards an Administrative Data Census after 2021. The full assessment can be found [here](#) [<<insert link when available>>](#).