

# Administrative Data – General Background Paper

## Why we're doing this?

Scotland's Census 2021 Programme specifies the following key aims relating to administrative data:

1. National Records of Scotland (NRS) plan to carry out a census in 2021 which is predominately online but which should make the best use of admin data where possible.
2. The scope of this census includes continuing to explore the future greater use of admin data in collecting socio-demographic statistics.
3. There may be an opportunity to provide enhanced outputs through the use of admin data and this will be explored.
4. Make recommendations for future censuses in Scotland.

We researched the role of administrative data as part of our Beyond 2011 programme of work which finished in 2014. Since then we have further developed our plans and we would like to share these with you in order to get your feedback.

## What are the project's objectives?

As a first step we are aiming to see if we can produce population and household estimates in a more timely, accurate and cost effective manner using administrative data. The project also aims to help Scotland's Census 2021 where possible. The specific objectives of the project are:

- To publish population and household estimates from admin data only – and independently of a Census - from summer 2020. The aim will be to produce these on an annual basis and then compare how these perform against NRS's current population and household estimates<sup>1</sup>. The ultimate test will be the comparison with Scotland's Census 2021.
- To help with Scotland's Census 2021: the linking methods we develop will be used to link the Census Coverage Survey (CCS) to Scotland's Census 2021. This is a key component of overall work to adjust the census estimates<sup>2</sup>. Errors in matching the CCS and census directly feed through to errors in overall population estimates and so it is crucial that matching is as accurate as possible. Improved methods will mean reduced time spent on clerical review as part of the linkage process potentially meaning outputs will be available earlier and possibly be more accurate.

---

<sup>1</sup><https://www.nrscotland.gov.uk/statistics-and-data/statistics/statistics-by-theme>

<sup>2</sup> Every effort is made to ensure everyone is counted in a census. However, no census is perfect and some people are inevitably missed. This undercount does not usually occur uniformly across all geographical areas or evenly across sub-groups of the population. To adjust for people who have not completed a census return, an independent Census Coverage Survey is conducted.

- To help with Scotland's Census 2021: to test Census Under Enumeration methodologies. This is where we don't get a return and then use administrative data to partially complete them. We need to test if these methods are statistically robust, with a view to potentially using this to improve census estimates.
- To help with Scotland's Census 2021: to explore the potential for enhanced outputs for Scotland's Census 2021 – this is where information from other administrative data is combined with census data - for example adding in information on the number of rooms.

We will continue to publish the official mid-year population estimates and household estimates which are ultimately dependent on the latest census on an annual basis. These will remain the official population and household estimates. We will compare the administrative data estimates with these official estimates to test and understand how well administrative data can be used as an alternative.. The ultimate test will be a comparison with Scotland's Census 2021.

#### Who else is doing this?

The Office for National Statistics have published administrative data population estimates for covering England and Wales for 2016 and 2017<sup>3</sup>. They produce an annual assessment of their progress towards producing information that is collected by a 10-yearly census (on housing, households and people) from an Administrative Data Census post-2021<sup>4</sup>.

For 2021 Census, NISRA will focus on planning for a census which will be predominantly online, while offering alternative modes of completion where necessary. It also aims to make best use of technology and administrative data in its design, building on the online approach used successfully in the 2011 Census. It has produced a number of research papers use of administrative data and data linking.<sup>5</sup>

Internationally, New Zealand's census 2018 has included administrative data for the first time. The final 2018 Census dataset consists of 89 percent census responses and 11 percent admin enumerations. The careful use of administrative data has enabled Stats NZ to add real data about real people to the 2018 Census dataset where they were sure they should have been counted but they had not received a census form. The inclusion of admin enumerations also removed the need to add missing census records through statistical imputation processes.<sup>6</sup>

<sup>3</sup><https://www.ons.gov.uk/census/censustransformationprogramme/administrativedatacensusproject/administrativedatacensusresearchoutputs/sizeofthepopulation>

<sup>4</sup><https://www.ons.gov.uk/census/censustransformationprogramme/administrativedatacensusproject/administrativedatacensusannualassessments>

<sup>5</sup> <https://www.nisra.gov.uk/publications/future-provision-census-population-information-northern-ireland>

<sup>6</sup> <https://www.stats.govt.nz/methods/overview-of-statistical-methods-for-adding-admin-records-to-the-2018-census-dataset>

## How we're doing this

We are developing methods which allow datasets to be de-identified in isolation at NRS before linking. To do this we remove names, addresses and dates of birth and replace these with “hashed matchkeys”.

These “hashed matchkeys” replace the personal identifiable data in each dataset and provide the “glue” to link different datasets together. De-identified datasets will then be linked in the National Services Scotland Safe Haven.

## Public and Stakeholder Feedback

Public feedback was sought from the Publics Panel in September 2017. The Publics Panel<sup>7</sup> is a group of lay members who are able to feedback public views about our work. We can then take account of their views in our decision-making and governance processes, ensuring our work reflects the interests and values of members of the public.

Stakeholder engagement is on-going and includes Local Authorities, members of the Population and Migration Statistics Committee (Scotland) (PAMS) and Privacy Groups. This feedback will be used to inform our work and help address privacy considerations.

To date we have met with all external data suppliers, building relationships and ensuring data sharing agreements are in place, these are periodically reviewed to make sure that all legalisation changes (GDPR in May 2018) and personnel changes are made.

## What data sources are being used in this study?

- Electoral Registers
- Higher Education Student Data (HESA)
- Scottish Government School Pupil Census
- Further Education Student Data
- Health Activity Data
- Vital Events – Births, Deaths, Marriages & Civil Partnerships
- NHS Central Register
- Census 2011
- Census 2021
- NRS Geography Data
- Registers of Scotland (RoS) Residential Sales

With the exception of NRS Geography data, data sources contain the name, data of birth, postcode and possibly more detailed address information. Some additional information is also available in some sources. For example, RoS data includes

---

<sup>7</sup> <https://adn.ac.uk/public-engagement/supporting-research/public-panels/>

house price information and health activity data contains the date of last known interaction with the health service – but no information on the part of the health service used – simply the date.

NRS Geography data contains a standardised list of all properties in Scotland, which can be mapped to a range of geographies across Scotland.

### Privacy, Security and Legal Issues

We have been working closely with colleagues to ensure that all legal, privacy and security issues have been addressed. These include compliance with all aspects of the Data Protection Act 1998 (DPA 1998) and subsequently, the General Data Protection Regulation (GDPR 2018 ) and ensuring that the environment used to do this analysis meets the relevant standards and guidelines, as set by the National Cyber Security Centre.

Under the GDPR 2018, the Data Protection Impact Assessment (DPIA) for this project, Administrative Data Mid-year Population and Household Estimates Project has been published on the Census 2021 website.<sup>8</sup>

### Longer Term Strategy

Once we have begun producing population and household estimates, the next step will be to test whether some kind of coverage survey is required to correct for under-reporting in the administrative data.

In addition to existing legal gateways, the Digital Economy Act 2017 which received Royal Assent on 27 April 2017, makes provision for data-sharing. The Act provides a permissive legal gateway for data acquisition and an opportunity to access UK Government datasets.

If we are to move away from a traditional enumerated census, administrative data cannot provide all of the characteristics information currently provided by a census. Therefore a separate population survey of characteristics would be required. The cost of such a survey will be a key factor in deciding whether this approach is financially viable. A survey will not be able to provide the small area breakdowns currently provided by a census, but could provide information on a more timely basis.

Lindsay Bennison

Head of Administrative Data

National Records of Scotland | Room 2/3/1 Ladywell House | Edinburgh EH12 7TF

Email: [Lindsay.Bennison@nrscotland.gov.uk](mailto:Lindsay.Bennison@nrscotland.gov.uk)

09 October 2019

---

<sup>8</sup> <https://www.scotlandscensus.gov.uk/Privacy-2021>