

2021 Census

Use of Administrative Data

April 2019



Northern Ireland
Statistics and Research Agency

Gníomhaireacht Thuaisceart Éireann
um Staitisticí agus Taighde

1. Summary

After each census the Registrar General for Northern Ireland carries out a review of the full census operation and produces a General Report. In 2014 Census Office published a report on the Future Provision of Census of Population Information for Northern Ireland information.¹ This paper stated that:

NISRA will focus on planning for a census in 2021 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.

In the UK the Office for National Statistics (ONS) are currently developing an Administrative Data Census and plan to produce it alongside the 2021 Census². ONS has produced research outputs to demonstrate to users what an Administrative Census might look like. Similar related work is ongoing in Scotland by the National Records of Scotland (NRS) and in the Republic of Ireland by the Central Statistics Office (CSO), Ireland.

NISRA will continually engage with ONS, NRS and Central Statistics Office (CSO) along with other international colleagues to ensure we are aware of the developments in, and the advantages and disadvantages of using administrative data in the census.

This document sets out current thoughts on how administrative data will be used in the 2021 Census in Northern Ireland. The document provides a brief description of the planned administrative data infrastructure followed by detailed proposals on how they could be used in the 2021 Census.

¹ <https://www.nisra.gov.uk/publications/future-provision-census-population-information-northern-ireland>

² <https://www.ons.gov.uk/census/censustransformationprogramme/administrativedatacensusproject>

2. Administrative Data Infrastructure

The administrative data used in the Census are collected and processed under the Census Act 1969 and used for statistics purposes only. The confidentiality of all census and administrative data is of paramount importance. All data is held on accredited secure systems and appropriate Information Assurance protocols are in place.

Key to the integration of administrative data into census processes is understanding the baseline data infrastructure. This can be categorised into three main components:

- I. *Address Register* – An address database created from administrative data sources providing indicators of property function (domestic/non-domestic), occupancy and other characteristics that would assist in the field operations of a census.
- II. *Statistical Person Database* - A limited person level record level database of the Northern Ireland population containing basic demographic information from administrative data sources.
- III. *Aggregated area based measures available on NINIS* – Benchmark or comparator information to be used to guide and assess respondent data quality across the Census process.

3. Administrative Data Integration

Current thoughts on the use of administrative data throughout the Census project have been identified for the following work areas or processing projects:

3.1 Operational

Address Register for use in enumeration

Key to the enumeration element of the census is the need for the field management and operations staff to have an accurate list of all addresses. The address register will be used to determine what addresses will receive invitations to complete a census return as well as how best to follow up on non-response.

The address register will also be key to providing a statistical address spine for data processing. This will include allocation of administrative geographies and associated attributes for outputs and Quality Assurance processes,

managing address response quality and identifying and removing any duplicate records.

Measures of Hard to Reach (HtR) for use in enumeration

Along with aggregated area based information, the administrative data infrastructure will be used as a basis to define HtR measures to allow field operations to determine best allocation of resource to maximise efficiency without compromising quality. Investigations will be carried out into incorporating Houses of Multiple Occupancy (HMOs), elderly person households, households without broadband/4G access and student information to aid enumeration.

3.2 Data Processing

Internal - Collection system coding frames

Depending on the electronic questionnaire (eQ) design, the administrative data may be used to auto-fill picklist options for address information. This may take the shape of valid addresses to complete the address on the front of the form, address 1 year ago, employer name and workplace address etc...

Internal - Demographic Data Cleansing

Lessons learned from the 2011 Census indicated that it would be of significant benefit to have cleansed demographic information (name, date of birth etc...) as early in the process as possible. Administrative data can aid this process by using name, address and date of birth lists as comparators to enumerated information to highlight records to be further assessed.

Measures of Hard-to-Count (HtC) for use in coverage assessment

Alongside the production of HtR measures will also be the derivation of HtC indicators. These are used during Data Processing to model, estimate and adjust the census returns for under and over enumeration. The address register, incorporating HtC indices, will also be used to inform the identification and selection of optimum Rehearsal and Census Coverage Survey (CCS) areas.

Coverage Assessment/Adjustment

Given the success of the Census Under-enumeration (CUE) project³ in the 2011 Census a similar project will be used in the 2021 Census. With further refinement and advancements in administrative data, it is expected that a CUE project will play a larger role in the 2021 Census. Investigations will be carried out into coverage modelling using administrative data sources along with a CCS.

³ <https://www.nisra.gov.uk/statistics/2011-census/results/ireland-and-northern-ireland>

Aggregated information from the administrative data infrastructure will be produced in 2021, alongside 2011-rolled forward mid-year estimates to enable users to assess the 2021 Census results. The final Census Estimate will be derived from the census enumeration adjusted by a coverage survey.

Item Imputation

Edit and Imputation processes were carried out in the 2001 and 2011 Censuses to counteract individual question non-response or inconsistencies. This process adopts a donor imputation methodology whereby a response value is identified by statistically identifying a best fit donor based on other record values. Whilst this method is statistically sound it is worth considering that administrative data may provide ‘real’ values and hence increase record level data quality.

Linking previous census information could provide information on religion, country of birth etc... Administrative data could provide information like household structure (number of adults, country of birth) and age of other HH members.

Linking to administrative data can be used to validate and quality assure census questions. For example the School Census data can verify if a child is in full-time education.

3.3 Outputs

Creation of new small area geographies

As part of the 2021 Census it may be a requirement to produce a new small area statistical geography. Aggregated household and person counts from administrative data could be used by to help design and create these prior to, or during, the 2021 Census process.

Replacing Questions

Census Office are continually balancing the user need for information along with the respondent burden and the demand for space on the Census form. Where there is continued user need for a question on the census but space is limited, and there is a good quality administrative source with the same information, an administrative data option will be explored. If found suitable the administrative data may be attached to the census and the information treated as an Enhanced Census Variable and the question will not be asked. Current examples include:

- Total number of rooms within a household
- Total number of bedrooms within a household

Quality Assurance

In the 2011 Census, a range of aggregated administrative data were used as quality assurance comparators to assess the census data as it progressed through Data Processing. It is expected that the role of administrative data in this capacity will increase. There may also be benefits in using record level address or person data to assist quality assurance.

Census Office

April 2019