

Quality Assurance for the 2019 Census Rehearsal

Quality Assurance of Administrative Datasets: Census Address Register

Dataset: Census Address Register - Rehearsal Version v4.1

Data supplier:	National Records of Scotland (Geography Branch)
Supplier info:	<p>National Records of Scotland (NRS) is a Non Ministerial Office of the Scottish Government. The purpose of NRS is to collect, preserve and produce information about Scotland's people and history and make it available to inform current and future generations.</p> <p>Within NRS the Geography branch provides geographic data and services that support the production of high quality statistics.</p>
Data content:	<p>The Census Address Register (CAR) is made up of a number of tables which contain:</p> <ul style="list-style-type: none">• Scottish address information• postcode information• census specific indicators and two unique identifiers;• the Royal Mail Unique Delivery Point Reference Number (UDPRN)• the One Scotland Gazetteer (OSG) Unique Property Reference Number (UPRN). <p>The CAR holds information on both current and historical addresses and postcodes and will be used throughout the census programme; from enumeration to outputs.</p>
Time period covered	Released on 30 August 2019 based on July 2019 data
Coverage:	The rehearsal covered 72,000 households across parts of Glasgow City, Dumfries and Galloway and Na h-Eileanan Siar. However the CAR included addresses from all of Scotland.
Use of data:	To test and quality assure methodology and processes in the 2019 Census Rehearsal

Data Source Information

The main input sources for the Census Address Directory (CAR) are the Scottish Address Directory (SAD) and the Scottish Postcode Directory (SPD). These input sources are described below. Additional information on communal establishments and Census specific information such as fieldwork areas, planning areas and enumeration types is also added. [Annex A](#) shows the full list of sources that are used to create the CAR.

Scottish Address Directory (SAD)¹

The SAD is used to populate over 90% of the address tables within the CAR. The SAD holds address information, postcode, council area, classification of type and unique reference identifiers for both possible residential and business addresses. It is created by combining elements from the data sources above, based on criteria set by NRS which utilises certain indicator values within the sources. This allows live and relevant addresses to be extracted and used in the SAD.

Key data sources used for the SAD are:

- [Ordnance Survey AddressBase Premium \(AB\)²](#) - an Ordnance Survey product which combines the One Scotland Gazetteer (OSG) with matched address information from the Royal Mail Postcode Address File (PAF).
- [Royal Mail Postcode Address File \(PAF\)³](#) - a database of all UK addresses and postcodes which receive a postal service from Royal Mail.
- [One Scotland Gazetteer \(OSG\)⁴](#) - Scottish address list combining the 32 corporate address gazetteers which are owned and maintained by each council. It contains address information and details of an addresses history as well as other information on type and location.

Scottish Postcode Directory⁵

The SPD is made up of two key datasets: Postcode Index - a file containing both live and deleted postcodes assigned to a variety of geographical areas in Scotland, and Postcode Boundaries - a file containing a digital boundary for every live small user postcode corresponding with the Postcode Index. NRS Postcode boundaries are based on the Postal Address File (PAF) received from Royal Mail. This data is provided on a quarterly basis. The files provide details of new and deleted postcodes. This information is augmented with advance planning information for new developments from local authorities. Using all available data, the Geography Branch digitisers plot the changes to the postcode dataset.

Data supply and communication

Members of the Geography branch visit local authority data custodians to build up a picture of issues with source data and develop a timeline of change. There are also

¹ <https://www.nrsotland.gov.uk/statistics-and-data/geography/our-products/scottish-address-directory>

² <https://www.ordnancesurvey.co.uk/business-government/products/addressbase-premium>

³ <https://www.poweredbypaf.com/>

⁴ <http://www.improvementservice.org.uk/one-scotland-gazetteer.html>

⁵ <https://www.nrsotland.gov.uk/statistics-and-data/geography/nrs-postcode-extract>

regular discussions between the Geography branch and with data suppliers including Ordnance Survey and the Improvement Service to highlight issues and find solutions that can be implemented at source.

The Geography branch participates in meetings of the Scottish Address Group every 6-8 weeks. This group includes colleagues from the Scottish Government, Improvement Service, Ordnance Survey, Geoplace and Scottish Assessors Association.

There are regular meetings between members of the Geography branch and Census programme, including an Address Quality Working Group, where the progress on any work can be discussed as well as any issues which may impact use of the address data.

Quality Assurance undertaken by data supplier

This section mostly gives details of the quality assurance performed by the NRS Geography branch. However the various products provided by external data suppliers are also subject to rigorous checks by these external suppliers. Information about products produced by external suppliers can be found on their websites using the links provided in the Data Source Information section on the previous page.

Scottish Address Directory (SAD)

As the SAD utilises data sources which were created and are maintained to meet the needs of external organisations there are anomalies within SAD. Some examples of these are:

- incorrect matches of address records between the OSG and PAF
- grid coordinates errors putting the building in the wrong position
- incorrect classification of the address (i.e. residential when it is in fact commercial)
- duplication of addresses
- vacant properties which are not occupied as residential or a business

There is an ongoing programme of quality assurance measures taking place to highlight and fix these anomalies where possible. These measures can be split into Quality Assurance measures and Quality Control measures. The Quality Assurance measures ensure that appropriate methods are being used to create products that meet the needs of their customers. The Quality Control measures are used to verify that the results of these methods are correct and working as expected. The measures used are listed below.

Quality Assurance:

- Internal cleaning of the source data to remove irrelevant information such as non-relevant addresses (e.g. fields, bus shelters) via a series of SQL queries.

- Comparison of address matches contained in AddressBase Premium and those from an independent address matching of PAF and OSG done by NRS Geography.
- Cross referencing address data between a variety of sources looking for inconsistencies.
- Comparing address classifications between different source datasets and investigating discrepancies.
- Regular discussions with Office of National Statistics (ONS) and Northern Ireland Statistics and Research Agency (NISRA) to share best practice and identify common challenges in production of Census address registers.
- Regular discussions with data suppliers including Ordnance Survey (OS), Geoplace and the Improvement Service (IS) to highlight issues and find solutions that can be implemented at source.
- Regular discussions with internal users of Geography address data.
- Visits to local authority data custodians to build up a picture of issues with source data and develop a timeline of change.

Quality Control:

- Regular production of management information statistics, which are monitored and analysed – this includes numbers of new, changed and deleted addresses, numbers of addresses meeting certain classifications and numbers of new, changed and deleted postcodes.
- Producer/review check i.e. work produced by one member of NRS Geography shall be checked by another member of NRS Geography before being released to customers.
- Targeted mail-out exercises to identify addresses where post is undeliverable.
- Field address check in areas of most concern e.g. areas where non-field address checks have a lower success rate.
- Peer review of address data by colleagues within NRS and Scottish Government.

Scottish Postcode Directory (SPD)

There are numerous checks to the each version of the SPD to ensure it is of high quality. These checks include:

- comparing each new version of the SPD to the previous version to ensure that no postcode content has been lost or become corrupt during production.
- running geography validation which covers relationship checks i.e. council area to health board area, council area to electoral ward, and all full extent geographies receive an allocation.
- checking that all live postcodes contain 1 grid reference and that there are no gaps (unless due to water) or overlaps on the postcode boundaries. Standard geometry checks are also run via ArcGIS.

Census Address Register (CAR)

Once a new version of the CAR has been produced a series of validation rules are run. These validation rules checks ensure that:

- variables are consistent across the different tables that make up the CAR
- particular variables, such as postcode, must be fully populated in certain tables.
- variables are unique within a table where necessary

Quality Assurance undertaken by the admin data team within the NRS Census Programme

The Admin Data team within the Census Programme had the task of formatting the address data so that it could be ingested by systems used in the 2019 Census Rehearsal.

As part of this work there was a check for duplication between the CAR and an address dataset for the rest of the UK (excluding Scotland) and 17 addresses were found to be duplicated. This information was passed back to the Geography branch so they could discuss with counterparts in the Office for National Statistics to resolve this issue.

There were also checks on the completeness of variables, consistency between historic and live addresses and that grid references for addresses made sense for the postcode provided. There were also checks that the Unique Property Reference Number for each address was unique.

Strengths and Limitations of the data source

Strengths	Limitations
<ul style="list-style-type: none"> • The main inputs, the SAD and SPD, are well established products and are updated on a regular basis. • The main inputs to the SAD and SPD themselves are commercial products that are subject to extensive quality assurance before being shared with the NRS Geography team. • The Census Address Register is produced specifically for use in the 2022 Census and the 2019 Census Rehearsal. • The CAR uses the same based addressing product as the Office of National Statistics (OS Addressbase premium). This provides benefits from harmonisation and cross-organisation work. • The Geography team have established excellent working relationships with their suppliers over a number of years. • The Geography team that create the CAR are viewed as subject matter experts. 	<ul style="list-style-type: none"> • The data needs to be a frozen in time to support census rehearsal. Therefore some new addresses will be missing, some addresses on the CAR will no longer exist and other addresses will change in some way. • There is a reliance on commercial products to create the CAR. As these products are created for another purpose they require some manipulation to make them fit for purpose for the census. • There could be small errors in the products used to create the CAR. These errors are not always easy to spot and will feed through into the CAR.

Risk/Profile Matrix

This section contains a risk/profile matrix for the NHSCR. The matrix reflects the levels of risk of data quality concerns and the public interest profile of the statistics. These have been determined by a review undertaken by the NRS Admin Data team using the information contained within the [Office for Statistics Regulation's Administrative Data Quality Assurance Toolkit](#). For the use of data for this project, the cell highlighted is appropriate:

Level of risk of quality concerns	Public interest profile		
	Low	Medium	High
Low	Statistics of low quality concern and low public interest. [A1]	Statistics of low quality concern and medium public interest. [A1/A2]	Statistics of low quality concern and high public interest. [A1/A2]
Medium	Statistics of medium data quality concern and low public interest. [A1/A2]	Statistics of medium quality concern and medium public interest. [A2]	Statistics of medium quality concern and high public interest. [A2/A3]
High	Statistics of high data quality concern and low public interest. [A1/A2/A3]	Statistics of high quality concern and medium public interest. [A3]	Statistics of high quality concern and high public interest. [A3]

*A1/A2/A3 – definitions supplied [Office for Statistics Regulation's Administrative Data Quality Assurance Toolkit](#).

Justification for Matrix Score

The public interest profile has been set to “medium” for the following reasons:

- There is a high level of interest in the Census and this work provides important information to assess the methods and processes to be used in the 2022 Census. Being able to assess these methods and processes accurately will therefore improve the quality of the 2022 Census which is used for a wide range of purposes, including planning, allocating funding and informing policy.
- However, as the Census Rehearsal does not produce outputs that will be used outside of improving Census methods and processes this is not of ‘high’ public interest.

The risk of quality concerns has been set to “low” for the following reasons:

- the programme of quality assurance measures means that errors in the data will be minimised.
- high level of communication between the Census programme and Geography branch to ensure the address data is being used appropriately and any issues can be discussed thoroughly.
- the Geography branch has well established relationships with their data suppliers to discuss the source data.
- this was the census rehearsal and any findings from this exercise will be used to make improvements for Scotland's Census 2022.

Annex A: Creating the Census Address Register (CAR) – input and processing map

