Paper Title: The Census Address Register

Presented by: Tricia Couper, Head of Geography, NRS

Security marking	Not sensitive
Can this be shared wider than PAMS?	Yes
Actions for PAMS members ahead of meeting	Paper is for information
Aims for discussion	
Key questions for PAMS members	
What is required from PAMS members	Whilst the paper is for information comments are invited from PAMS members.

Population And Migration Statistics (PAMS) Committee (Scotland)

Census Address Register - update

Introduction

1. Scotland's Census 2022 aims to encourage completion online and for follow up activity to be targeted, this approach is reliant on high quality address information to ensure we can make contact with every household in Scotland.

2. Our Geography team is responsible for creating a Census Address Register (CAR) that will be used to underpin census operations, several iterations of the CAR have been created for quality assurance purposes. The CAR is made up of a number of tables which contain Scottish address information, postcode information, census specific indicators and 2 unique identifies; the Royal Mail Unique Delivery Point Reference Number (UDPRN) and the One Scotland Gazetteer (OSG) Unique Property Reference Number (UPRN)

Source data

3. The CAR uses Ordnance Survey's (OS) AddressBase Premium (ABP) product which takes the local authority addresses from the One Scotland Gazetteer and matches them to the Royal Mail Postcode Address File (PAF).   Geography then attaches additional information from internal and external data sources. Holding the unique reference numbers allows address data to be linked to a number of other datasets which holds these identifiers.   This includes Scottish Assessors (SAA) data, Electoral Registration data, NRS Communal Establishment Register (CER) which gives further details of the type, residents and management of the communal establishment as well as other external datasets. Data from the SAA and ER data are key in determining whether the residential status of an address in the source ABP data has been correctly classified residential.

Quality

4. Errors in CAR can have a range of impacts to NRS. These could be reputational, as if we are seen to be using data which is inaccurate people may be less inclined to trust the NRS, or increased costs as if we include addresses for the census which are historical or not yet occupied this wastes time and materials and increases the likelihood of undelivered mail having to be dealt with.

5. The Geography team have developed quality assurance and quality control measures to ensure that appropriate methods are being used to create the products and verify that the results of the methods are correct and working as expected. Unfortunately a number of our planned visits to local authorities and physical ground checks have not taken place due to COVID restrictions.

Quality assurance

*   Internal cleaning of the source data to remove irrelevant information.
*   Comparison of address matches contained in ABP and those from independent matching of OSG to PAF by NRS Geography.

- Analysis of council tax and valuation roll address data to identify where unique identifies are missing or duplicated therein, include a matching exercise of this data to OSG.
- 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 ONS and NISRA to share best practice and identify common challenges.
- Regular discussions with data suppliers to highlight issues and find solutions that can be implemented at source.
- Regular discussions with internal users of Geography address data.

Quality control

- Regular production of management information statistics which are monitored and analysed over time – this includes the numbers of new, changed and deleted addresses, numbers of addresses meeting certain classifications, numbers of new changed and deleted postcodes and records holding conflicting information across multiple datasets.
- Peer review of CAR business rules. (These rules determine whether a record meets residential criteria)
- Producer/reviewer check.
- Targeted mail out exercise to identify where post is undeliverable.

I**mprovements in source data**

6. The Geography team have worked closely with the Improvement Service over the past few years and attended a number of OSG community meetings where they've been able to speak directly to the CAG custodians. Therefore we are aware of the improvements already taking place and have been lucky enough to provide input into these.

7. The main improvement areas for us are;

- classifications to at least a secondary level as this gives us greater confidence in identifying if an address is residential, a business or in some cases an unwanted record (for example ATM's or land),
- accurate grid co-ordinates as we use these to maintain our postcode boundaries and ensure we have grouped all the addresses correctly, critical for accurate census outputs and
- updated property lifecycle information as this allows us to remove provisional, unoccupied or not yet built properties from our datasets and monitor rates of change.

Quality assurance work

8. Below are some examples of recent quality assurance work:

Organisation check. We are checking around 5,000 records where there is an organisation name listed as part of the address, therefore implying a non-residential status but the record holds a residential classification in source ABP data or has a match to Council Tax.

Address Mail out. We are planning to run another mail out exercise which had been on hold until after the Parliamentary elections to verify the status of an address or check if the address has an alternative name.

Provisional records with a Royal Mail match. There are just over 800 records holding a provisional flag in source ABP (which means not yet approved by planning) but have been matched to a live Royal Mail address in source ABP. These records are being investigated to determine if:

- They are in use and therefor the local authority holds the incorrect information, or
- They have been wrongly added to PAF too early, or
- The address match between PAF and the local authority address is incorrect

We found that 15% had been matched incorrectly in the source data and 64% incorrectly flagged provisional instead of live. We suspect the latter may down to timing issues rather than error. Findings will be sent back to suppliers.

Under construction records with a Royal Mail match. There are approximately 17,000 records marked as being under construction by the local authority by have been given a match to a Royal Mail address by our supplier. The address has been compared to recent council tax data and around 11,600 of these records appear on the list. These addresses have been included in the postcode boundary amendments work to help identify any where there is an error with one of the data sources.

Matching. Independent matching of ABP matches (OSG:PAF), matching to Council Tax and Electoral Register data continues.

Cleaning Rules. Ongoing refinement of cleaning rules applied to remove unwanted records from the data. This has helped to build up greater understanding of the information available, what issues there may be and ultimately to help ensure the data being used is relevant and of the highest possible quality.

Communal Establishment Register (CER) Work to identify CE and CE sub addresses is ongoing, recently focus has been identifying groups of 3 or more self-catering lodges or cottages which will along with other CE be subject to an address check by NRS Census Enumeration team.

UPRN matrix. A benchmark has been created to pull together all the information from across the different address sources and the various quality assurance exercises which Geography are undertaking. It highlights the confidence level of addresses based on factors such as classification, match to Royal Mail and if they appear on the council tax list or the electoral register. Over 93% of the total records in the CAR (so that is residential and non-residential) have been matched to 3 or more sources, e.g. RM match which is verified by Geography, council tax agreement across 2 or 3 organisations (assessors, GeoPlace and NRS), residential classification on source, electoral register match, communal establishment. A further 6% matched to less than 3 sources and 1% need further analysis.

NRS: Geography
May 2021