

POPULATION AND MIGRATION STATISTICS COMMITTEE (SCOTLAND)

Census Output Geographies

Introduction

1. This paper seeks to provide some background and highlight some of the methodological issues related to geographies for the 2011 Census outputs. In particular it:

- Provides a recap of the set of higher level geographies for which standard 2001 Census outputs were produced;
- outlines the approach being developed to create census outputs areas for 2011;
- describes the current methodology used for defining settlement and locality boundaries;
- flags up a variety of methodological and other issues.

Background

2. The results from the 2001 Census were made available for a variety of geographies. The building bricks for all higher geographies were census output areas (OA) which were created as groups of postcodes nesting as well as possible into (in descending order of preference) the following areas: Council Area, 2001 locality, 1991 OA, postcode sector and 2001 electoral ward. On average each OA contained around 50 households (with minimum thresholds for confidentiality set at 20 households and 50 individuals). A “master postcode” for each OA was determined, with the OA inheriting all the characteristics of that postcode, including its assignment to higher geographies* and its centroid grid reference.

3. The higher geographies for which standard census outputs were generated included:

- Council areas
- Westminster and Scottish Parliamentary constituencies
- Health Board areas
- Postcode sectors**
- 2001 wards**
- Settlements and localities
- Civil parishes
- Inhabited islands
- Ad hoc areas (users were able to generate outputs for their own ad hoc combinations of OAs.)

Census counts of the number of households and residents in each postcode were also generated during processing.

* OAs aggregate exactly to council areas but not necessarily to other higher geographies (where they will aggregate as best fit approximations).

** Two types created, to reflect the different confidentiality thresholds that applied to outputs for Census Area Statistics and to Standard Tables.

2011 Census Output Areas

4. As in 2001, it is proposed to create census output areas (OAs) from postcodes to form the building bricks for census outputs for all higher geographies. The main aim when creating OAs will be to ensure as much continuity as possible with OAs from previous censuses. The approach will therefore be to allocate all 2011 postcodes to 2001 OAs and then merge OAs (where they have dipped below confidentiality thresholds) or to split them (e.g. when new housing has taken them above maximum thresholds). Changes since 2001 in locality boundaries will also be taken into account where possible, as well as the opportunity to balance OAs by moving postcodes (though not at the expense of continuity).

5. In creating OAs for 2011 it is planned to adopt the following basic principles:

- Continuity – OAs in 2011 to be as similar to 2001 as possible
- Minimum thresholds – 20 households and 50 individuals (same as in 2001)
- Postcode content of OAs – nesting (in order of preference) to be council area, 2011 locality, 2001 OA
- OAs and postcodes – all postcodes in an OA to belong to same locality (where possible)
- OA target population and shape – aim for target size of 52 households (with residents) and best shape as circle (if that were possible)

6. Further details of the proposed methodology for creating 2011 Census output areas is given in the Annex to this paper.

Settlement/ locality boundaries

7. A settlement is currently defined as a contiguous group of “high density” postcodes (bounded by “low density” postcodes or water) with a population of 500 or more. A postcode is defined as “high density” (i.e. urban) if one or more of the following three threshold conditions are met:

- The number of residential addresses per hectare exceeds 2.1
- The population per hectare exceeds 5
- The number of non-residential addresses per hectare exceeds 0.1

Otherwise, the postcode is classified as “low density” (i.e. rural).

8. Localities are intended to be more representative of the towns and cities in Scotland. Some settlements cover an extensive area, and consist of more than one distinct town or city. For example, the settlement of Glasgow consists of many towns, such as Paisley, as well as the city of Glasgow itself. So, some settlements are divided into constituent towns or cities (‘localities’). The 2001 Census report ‘Key Statistics for Settlements and Localities Scotland’ is used as the basis for defining localities.

The current intention is to apply the definitions and thresholds mentioned above in updating settlement and locality boundaries using the 2011 Census dataset.

9. A number of issues have been identified in connection with settlements and localities, including:

- the handling of postcodes which straddle both the edge of a settlement and a rural hinterland
- the relevance of settlements and localities in relation to things like service delivery
- whether the thresholds for density of postcode are set at the right levels, and whether different thresholds should apply in different parts of Scotland, e.g. to take account of crofting communities.

Other issues

10. *Scottish Neighbourhood Statistics*

At the time of the 2001 Census, the use of information from Scottish Neighbourhood Statistics was not as well established as it now is, and no standard census outputs were generated for SNS data zones. However, given the importance of SNS to many users, particularly in local authorities, it is planned that SNS data zones will be included as one of the geographies for which standard 2011 Census outputs will be produced. Data zones are built from aggregations of census output areas, and so any changes to the latter will impact on the former. The creation of census output areas for 2011 will also provide an opportunity to realign data zone boundaries where appropriate, though time series requirements and how data zones fit to higher level geographies will need to be considered.

11. *Frozen geography*

In 2001 it was decided that the small user postcodes current at the time Enumeration Districts were finalised ahead of enumeration should be the standard basis for all census output geographies. Any postcodes encountered during processing that were not in the frozen set were replaced by ones that were. This was seen to provide many advantages: all geographies (enumeration, travel destination, migration origin) were put on the same postcode basis; creating OAs out of postcode of enumeration automatically created OAs out of the other geographies; users were aware of the postcode basis well in advance of receiving any output. These all appear to be strong arguments in favour of following a similar approach for 2011.

12. *Standard codings for geographic areas*

Scottish Government and GROS is currently considering how a draft National Statistics statistical geography coding and naming policy from ONS will apply in Scotland. Full implementation of the policy is scheduled for 2011, when the outcome of current reviews of Parliamentary Constituencies and administrative boundaries will also be taken into account. One potential issue relates to the extent to which meaningful labelling should be associated with the standard codings, e.g. for SNS data zones.

13. *Non-contiguous postcode boundaries*

As the geographies used by GROS for census outputs are based on postcodes, they can be affected by anomalies created by non-contiguous postcodes, i.e. postcodes that have two or more discrete parts. This can arise for example when a new housing development might place new addresses within the area of an existing postcode (thereby splitting it), or when Royal Mail allocates the same postcode to addresses on the mainland and also to addresses on a nearby island. The extent of this issue is not huge – it affects only 521 postcodes (out of a total of approximately 145,000) current small user postcodes. These postcodes contain 9,914 addresses, 83% of which fall within the parts of the individual postcodes which contain most the addresses.

14. ***Workplace output areas***

There has been interest from some quarters in the possibility of census results being generated for a “workplace zone” geography, i.e. based on the distribution of workplaces rather than households. While this has some attractions, weighed against it is the relatively low data quality obtained for the workplace address question in previous censuses and the fact that the census is not designed as a census of workplaces.

15. ***User consultation on area boundaries (e.g. localities, SNS data zones)***

The processes used to create boundaries for census output areas, localities and SNS data zones are largely automated, making it feasible and relatively inexpensive to maintain and update them. Manual adjustments are time-consuming to make and a potentially error-prone process. Previous reviews and the ongoing dialogue that GROS Geography has with local users about boundaries should also mean the current boundaries largely reflect user needs. However, it is recognised that some users may want to provide specific input into the design of the 2011 Census localities and to any realignment of SNS data zones.

16. ***Other issues***

Choices still to be made on the standard population base(s) for statistical outputs from the 2011 Census and on the method of statistical disclosure control to be applied have the potential to impact on census output geographies.

Further work

17. GROS plans to hold a further formal round of consultation on the 2011 Census towards the end of 2009 (by which time the questions to be asked in the census, the population definitions involved and the methodology for statistical disclosure control should be more or less settled). That consultation will have census outputs as its main focus, and will certainly include geography as a specific topic on which views will be sought. However, given the detail and complexity of some of the geographical issues involved it may be worthwhile exploring them separately, perhaps via a time-limited sub-group of PAMS.

Views and comments of PAMS members on the various issues mentioned above are invited.

Annex: 2011 Census Output Areas

1. Background

Output Areas (OAs) are the lowest level of Geography at which statistics are released in Scotland and are created by grouping postcodes together. All other statistical output area types (eg Settlement) are constructed from building up OAs although not all are an exact fit. The main aim when creating OAs is to ensure that there is continuity with OAs from previous censuses, whilst also making sure that OAs respect changing Locality boundaries where possible.

Due to a change in software tools available from 2001, the starting point for creating OAs is different for 2011. We will begin by allocating all the 2011 postcodes to 2001 OAs (which will be done by using grid references). This will give us a starting point from which to begin constructing OAs for 2011. There will be roughly 42,604 OAs at this stage (the total number of 2001 OAs). Some of the OAs will fall below the minimum thresholds due to houses having been demolished since 2001 and some will be above the maximum thresholds due to new houses being built. There is also the issue of changing Locality boundaries to contend with. Some boundaries may have changed due to, for example, new houses being constructed on the edge of settlements.

It is proposed that OAs for 2011 will be created as groups of postcodes nestling as well as possible into the following areas: Council Area (CA), 2011 Locality Code and 2001 OA. These areas are listed in order of preference, which has to be applied when postcodes don't fit into a single combination of these. All OAs must fit within a CA and it is desirable that they also fit within the other areas. Postcode Sector and Electoral Ward were taken into consideration in 2001 and, due to the proposed methods of splitting and merging of OAs (mentioned below), there will remain a good level of continuity with these two areas. Postcode Sector and Electoral Ward will therefore not be taken into consideration directly in the Business Rules used to create 2011 OAs.

If an OA is over the maximum threshold then it should, where possible, be split to create two OAs. If the OA is vastly over the threshold then it may be split more than once. When creating OAs, a size of 52 households (2001 average) and 119 people (2001 average) will be targeted as well as trying to have an OA in as compact a shape as possible.

We will try to balance OAs in order to improve them. This will be done by moving a postcode from one OA to another. The factors involved in balancing will be weighted and if moving a postcode improves the score of the OAs involved then it will be moved. The criteria to be filled when moving a postcode between OAs are: Postcode belongs to the same CA/Locality, Postcode in same OA as in 2001, Shape as close to circle as possible and Size as close to 2001 average as possible.

Where OAs are too small we plan to merge these with a neighbouring OA. The first choice would be to merge the OA that is too small with the 2001 OA which it belonged to. This may occur when a 2001 OA has been split and has resulted in a 2011 OA which now falls below the minimum threshold. Where this is not possible the OA will be merged with the smallest neighbouring OA.

The Business Rules for creating 2011 OAs are outlined in the following section.

2. 2011 Business Rules

1. Continuity

- a) OAs in 2011 should be as similar to 2001 as possible.

2. Minimum OA criteria (for confidentiality).

- a) Households (HH) 20
- b) Population (Pop) 50

3. Postcode content of OAs

Postcodes should nest into the following areas (listed in order of preference, which has to be applied when postcodes don't fit into a single combination of all of these).

- a) Council Area (CA)
- b) 2011 Locality Code
- c) 2001 OA

4. OAs and Postcodes

- a) Postcodes that straddle CAs are split by GRO and so they will be split between OAs.
- b) All postcodes in an OA must belong to the same CA.
- c) Where possible all postcodes in an OA should belong to the same Locality (desirable but not always achievable).
- d) Conversely, rural postcodes to be grouped into 'rural' OAs.

5. OA's target population and shape

- a) Aim for a target size of HH 52 and Pop 119 (2001 average).
- b) Best shape would be a circle (if that were possible).

6. Splitting OAs

- a) If an OA has HH > 104 and Pop > 238 then try to create two OAs.
- b) If an OA has HH > 156 and Pop > 357 then try to create three OAs.
- c) and so on etc.
- d) If a Locality postcode is in an old 2001 rural OA then make this postcode an OA on its own (later merge with a suitable OA in the appropriate Locality).

7. Balancing OAs

Move postcodes between suitable OAs to achieve target HH and Pop and a good shape.

- a) Postcodes in the OA to belong to the same CA/Locality.
- b) Postcodes (determined by grid reference position and 2001 OA polygon) to be in the same OA as in 2001.
- c) OA size to be as close to 2001 average as possible.
- d) Shape as close to circle as possible.

8. Merging OAs

Up to this point, all the postcodes in an OA should have the same 2001 OA code. The merging process will potentially break this rule. This will be due to geography or a changed (from 2001) Locality boundary or housing demolition.

- a) Small OAs are to be merged with the best candidate from neighbouring OAs in the same CA.
- b) The OAs should have same 2001 OA code.
- c) The OA should also belong to the same type ie. a rural OA should not be merged with a locality OA and vice versa).
- d) If b) or c) is not possible then merge with the smallest neighbouring OA.
- e) Where a choice is available the best shape (of merged OAs) should be chosen.