Evaluation of the Data Zone Geography

Monitoring Population Drift

February 2011

Scottish Government
Geographic Information Science & Analysis Team (GI-SAT)
Table of Contents

1. Introduction ........................................................................................................................................3
2. Demographic Changes ......................................................................................................................5
3. Boundary Changes ...........................................................................................................................11
4. Conclusions ......................................................................................................................................23

Figures and Tables

Figure 1.1: Largest (left) and smallest (right) data zones in terms of area extent. .............................................................. 3
Table 2.1: GROS Small Area Population Estimates, mid-year estimates 2001 - 2009 ............................................................ 5
Figure 2.1: Percent of data zones out-with the 500 - 1000 population range, 2001 - 2009 ................................. 6
Figure 2.2: Data zone population distribution, 2001 and 2009 ...................................................................................... 7
Figure 3.1: Data zone crossed by Local Authority boundary, one postcode shown is classified inaccurately. ........ 11
Table 3.1: Total number and percentage of postcode units that are classified inaccurately, for seven geographies. .. 12
Figure 3.2: Total number of postcode units that are classified inaccurately, for seven geographies. .......... 13
Figure 3.3: Percent of postcode units that are classified inaccurately, for seven geographies .............................. 14

Maps

Map 2.1: Population per data zone in 2001 .............................................................................................................................. 8
Map 2.2: Population per data zone in 2009, SAPE mid-year estimate ................................................................. 9
Map 2.3: Population trend by data zone, 2001-2009 ............................................................................................................ 10
Map 3.1: Postcodes assigned to the wrong Urban/Rural category when using data zones to group postcode units. . 16
Map 3.2: Postcodes assigned to the wrong Health Board when using data zones to group postcode units ........... 17
Map 3.3: Postcodes assigned to the wrong Local Authority when using data zones to group postcode units .......... 18
Map 3.4: Postcodes assigned to the wrong Community Health Partnership when using data zones to group postcode units. ........................................................................................................... 19
Map 3.5: Postcodes assigned to the wrong Westminster Parliamentary Constituency when using data zones to group postcode units. ........................................................................................................... 20
Map 3.6: Postcodes assigned to the wrong Scottish Parliamentary Constituency when using data zones to group postcode units. ........................................................................................................... 21
Map 3.7: Postcodes assigned to the wrong Multi Member Ward when using data zones to group postcode units. .... 22
1. Introduction

Scotland is divided into 6,505 data zones of varying size and shape, but with a roughly standard population in both rural and urban areas. The creation and definition of data zones was based on years of consultation with users of geographical statistics in Scotland. The final publication of the data zone geography and related statistics through Scottish Neighbourhood Statistics (SNS) in spring 2004 was seen as a milestone in the ability to monitor and develop policy at small area level, for the Scottish Executive (now Scottish Government) and partners. Data zones are increasingly becoming the core geography for small area statistics across most policy areas, such as education, health and the labour market. Full background information on data zones and the methodology used to create them is available online or as a pdf.

The main criteria used in the initial definition of data zones in approximate order of importance:

1. Approximate equality of resident population (500-1000 persons)
2. Compactness of shape
3. Approximate homogeneity of social composition
4. Existence, where possible, of some community of interest
5. Accordance with other boundaries of local significance
6. Accordance with prominent features in the physical environment

During the initial creation of the data zones it was acknowledged that often these constrains could not all be met. Therefore, a good compromise between these constrains should be found for each data zone. The smallest data zone, located in Edinburgh (Figure 1.1, right), covers an area of only 12,367 square metres and has a population of 496. The largest data zone is located in the Highlands and covers an area of 1,159 square kilometres with a population of 689 people.

Figure 1.1: Largest (left) and smallest (right) data zones in terms of area extent.

Data zones are meant to be used as static geographic units over time. However, three issues related to temporal changes, affect their validity as such:

1) **Demographic Changes:** An increasing number of data zones have a population of more than 1,000 or less than 500, due to population changes over time.

2) **Boundary Issues:** Data zones are frequently split by boundaries of other geographies, leading to inaccurate classifications of output areas and postcode units, due to two main issues:
   a) Geographies can be derived from a variety of building blocks, such as parishes or postcode units, which are non-coterminous with each other.
   b) Geographies may have been coterminous with data zones in 2001, but are not static and may have since been modified.

3) **Loss of Social Homogeneity:** By definition, the data zones were created to have social homogeneity. However, over time social and economic changes will have taken place that may not be homogeneous within one data zone.

The purpose of this report is to evaluate the effects of these issues on data zones as a stable geographic unit. Section 2 looks at the changes in populations of data zones since 2001 using yearly estimates published by the General Register Office for Scotland (GROS). The changes in the distribution of population are observed to monitor the increasing number of data zones falling out-with the 500 – 1000 population group.

Section 3 looks at the data zone geography’s accordance with seven other commonly used boundaries. This level of fit is quantified by calculating the number of postcodes that are inaccurately classified by data zones when aggregated to these other geographies. Each year, any changes to the boundaries of these geographies are noted and the results are recalculated.

Table 1.1 shows the datasets used in this evaluation. As of yet, the loss of social homogeneity within data zones has not yet been dealt with, but it is expected that future versions of this evaluation will incorporate some measure of this issue.

**Table 1.1: Datasets used for this report**

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Census 2001 Household Population Figures³</td>
<td>General Register Office for Scotland (GROS), 2001</td>
</tr>
<tr>
<td>Small Area Population Estimates (SAPE)⁴</td>
<td>General Register Office for Scotland (GROS), 2001-2009</td>
</tr>
<tr>
<td>Postcode Index, Release 2</td>
<td>General Register Office for Scotland (GROS), 2010</td>
</tr>
<tr>
<td>Health Board Area</td>
<td>Scottish Executive, 2006</td>
</tr>
<tr>
<td>Local Authority</td>
<td>Ordnance Survey (OS) BoundaryLine, 2010</td>
</tr>
<tr>
<td>Community Health Partnership</td>
<td>Scottish Government, 2007</td>
</tr>
<tr>
<td>Westminster Parliamentary Constituency</td>
<td>Ordnance Survey (OS) BoundaryLine, 2005</td>
</tr>
<tr>
<td>Scottish Parliamentary Constituency</td>
<td>Boundary Commission, 2010</td>
</tr>
<tr>
<td>Multi Member Wards</td>
<td>Ordnance Survey (OS) BoundaryLine, 2010</td>
</tr>
<tr>
<td>Urban/Rural Classification (6 and 8-fold)</td>
<td>Scottish Government, 2009-2010</td>
</tr>
</tbody>
</table>

³ [http://www.scrol.gov.uk/scrol/common/home.jsp](http://www.scrol.gov.uk/scrol/common/home.jsp)
2. Demographic Changes

To enable the use of data zones as a static geography, the population living in each data zone should be within a certain range for two main reasons. Firstly, to compare population-related statistics for certain areas in Scotland, it is essential that the number of people living in each geographical unit is both comparable with other data zones and relatively stable over time. Secondly, if the number of persons living in a particular data zone is much less than the average, disclosure problems and outliers are more likely to occur.

As described in the methodology used to create the data zones, initially 3.8% of the data zones were outside the 500 - 1000 population range. The outlier data zones were accepted because the other criteria, mentioned in the introduction were largely met. The population figure used to determine the data zone boundaries was the Census 2001 household population count, which does not include people living in communal establishments. Based on the total population count, initially 5.9% of data zones were outside the 500 - 1000 population range.

GROS publishes mid-year small area population estimates (SAPE), usually in the autumn of the following year (i.e. Mid-year estimates for 2009 were published in September 2010). The mid-year estimates for each of the consecutive years is a function of the Census 2001 total population count, the migration, death, birth, student and armed forces figures. Based on SAPE data 6.2% of data zones were outside the original 500 - 1000 population range on 30 June 2001, nine weeks after the Census 2001 population count. This percentage has continued to grow over the years to 14.3% in 2009. Table 2.1 shows the population per data zone divided into five population classes for each year from 2001 to 2009.

Table 2.1: GROS Small Area Population Estimates, mid-year estimates 2001 - 2009

<table>
<thead>
<tr>
<th>Year</th>
<th>&lt; 300</th>
<th>300 - 499</th>
<th>500 - 999</th>
<th>1000 - 1499</th>
<th>1500 +</th>
</tr>
</thead>
<tbody>
<tr>
<td>2001</td>
<td>0</td>
<td>29</td>
<td>6,104</td>
<td>360</td>
<td>12</td>
</tr>
<tr>
<td>2002</td>
<td>0</td>
<td>65</td>
<td>6,044</td>
<td>384</td>
<td>12</td>
</tr>
<tr>
<td>2003</td>
<td>0</td>
<td>88</td>
<td>5,960</td>
<td>444</td>
<td>13</td>
</tr>
<tr>
<td>2004</td>
<td>1</td>
<td>109</td>
<td>5,883</td>
<td>495</td>
<td>17</td>
</tr>
<tr>
<td>2005</td>
<td>2</td>
<td>113</td>
<td>5,815</td>
<td>552</td>
<td>23</td>
</tr>
<tr>
<td>2006</td>
<td>2</td>
<td>129</td>
<td>5,751</td>
<td>584</td>
<td>39</td>
</tr>
<tr>
<td>2007</td>
<td>4</td>
<td>127</td>
<td>5,714</td>
<td>600</td>
<td>60</td>
</tr>
<tr>
<td>2008</td>
<td>4</td>
<td>138</td>
<td>5,638</td>
<td>649</td>
<td>76</td>
</tr>
<tr>
<td>2009</td>
<td>7</td>
<td>159</td>
<td>5,572</td>
<td>685</td>
<td>82</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Year</th>
<th>No. of Data Zones</th>
<th>Minimum Population</th>
<th>Maximum Population</th>
<th>Mean</th>
<th>Median</th>
<th>5th Percentile</th>
<th>Lower Quartile</th>
<th>Upper Quartile</th>
<th>95th Percentile</th>
</tr>
</thead>
<tbody>
<tr>
<td>2001</td>
<td>6,505</td>
<td>477</td>
<td>2,815</td>
<td>779</td>
<td>775</td>
<td>546</td>
<td>666</td>
<td>886</td>
<td>1,006</td>
</tr>
<tr>
<td>2002</td>
<td>6,505</td>
<td>453</td>
<td>2,859</td>
<td>777</td>
<td>771</td>
<td>547</td>
<td>663</td>
<td>885</td>
<td>1,012</td>
</tr>
<tr>
<td>2003</td>
<td>6,505</td>
<td>377</td>
<td>2,841</td>
<td>777</td>
<td>769</td>
<td>543</td>
<td>662</td>
<td>885</td>
<td>1,020</td>
</tr>
<tr>
<td>2004</td>
<td>6,505</td>
<td>248</td>
<td>3,200</td>
<td>781</td>
<td>769</td>
<td>540</td>
<td>661</td>
<td>888</td>
<td>1,036</td>
</tr>
<tr>
<td>2005</td>
<td>6,505</td>
<td>224</td>
<td>4,024</td>
<td>783</td>
<td>768</td>
<td>537</td>
<td>660</td>
<td>888</td>
<td>1,049</td>
</tr>
<tr>
<td>2006</td>
<td>6,505</td>
<td>0</td>
<td>4,510</td>
<td>787</td>
<td>768</td>
<td>537</td>
<td>657</td>
<td>888</td>
<td>1,071</td>
</tr>
<tr>
<td>2007</td>
<td>6,505</td>
<td>0</td>
<td>5,219</td>
<td>791</td>
<td>767</td>
<td>536</td>
<td>657</td>
<td>891</td>
<td>1,091</td>
</tr>
<tr>
<td>2008</td>
<td>6,505</td>
<td>0</td>
<td>6,453</td>
<td>795</td>
<td>768</td>
<td>535</td>
<td>658</td>
<td>894</td>
<td>1,118</td>
</tr>
<tr>
<td>2009</td>
<td>6,505</td>
<td>0</td>
<td>7,061</td>
<td>798</td>
<td>767</td>
<td>531</td>
<td>658</td>
<td>896</td>
<td>1,137</td>
</tr>
</tbody>
</table>
In 2001 93.8% of the data zones were within the 500 – 1000 population range. A total of 29 data zones had a population of less than 500, but none less than 300. A total of 372 data zones had a population of 1000 or more, with 12 having had a population of more than 1500. The maximum population in a single data zone in 2001 was 2,815.

By 2009 the percentage of data zones within the 500 – 1000 population range has dropped to 85.7% (a further 8.1%). Therefore, over 14% or 933 data zones were out-with the 500 - 1000 population range in 2009. The number of data zones, with a population of either less than 300 or more than 1500, has increased from 12 to 89 since 2001.

Of the five population size classes mentioned in Table 2.1, four are out-with the 500 - 1000 population range. The graph in Figure 2.1 shows the percentage of data zones in these four classes over the years 2001-2009.

![Figure 2.1: Percent of data zones out-with the 500 - 1000 population range, 2001 - 2009](image)

From Figure 2.1 it appears that the number of data zones with a population below 500 has grown gradually between 2001 and 2009. In addition, the number of data zones with a population greater than 1000 has increased substantially between 2001 and 2009. Based on this trend it can be expected that the total number of data zones out-with the 500 - 1000 population range will continue to grow each year – although some individual data zones may now have reached their capacity in terms of new build.

Figure 2.2 shows the population size distribution for both 2001 and 2009. From these graphs a gradual lowering and widening of the distribution curve is observed from the period of 2001 to 2009.
Maps 2.1 and 2.2 show thematic maps of the distribution of the five population classes in 2001 and 2009 respectively. Map 2.3 shows the data zones that have exhibited a continuous population growth or decline from 2001 to 2009. Red areas have consistently increased in population, whereas green areas have consistently decreased in population. Darker shades of red and green indicate whether the data zone, as of 2009, is out-with the 500 – 1000 population change.

Figure 2.2: Data zone population distribution, 2001 and 2009
Map 2.1: Population per data zone in 2001
Map 2.2: Population per data zone in 2009, SAPE mid-year estimate.
3. Boundary Changes

Data zones are used by Scottish Neighbourhood Statistics (SNS) to link information to other (larger) geographies. Therefore, one data zone can not be located in more than one area of any geography, such as a Local Authority or Scottish Parliamentary Constituency, even if the boundaries cross. SNS provides values for several geographies for each data zone based on either the location of the data zone centroid or the location of the majority of the Census output areas that make up a data zone.

Boundaries of other geographies, such as Local Authorities or Scottish Parliamentary Constituencies are not fully coterminous with data zone boundaries and a data zone can be split in two or more parts by these boundaries (see Figure 3.1). The two main issues causing this mismatch are the use of different building blocks such as postcodes or parishes which are non-coterminous, or changes to the boundaries that have occurred since 2001, when data zones were first created.

During the initial creation of data zones, one of the conditions was the accordance of data zone boundaries with other boundaries of local significance. In this section this accordance is tested to seven often used Scotland-wide geographies.

![Figure 3.1: Data zone crossed by Local Authority boundary, one postcode shown is classified inaccurately.](image)

The indicator used to quantify the effect of boundary issues on the data zone geography is the number of inaccurately classified postcodes, based on two different lookup methods. The smallest geography used in this analysis is the Postcode Index (PCI). When this evaluation was
initially completed in 2006, the 2006 (release 2) of the PCI was used, however, for subsequent years when boundaries have been amended, the most recent version of the PCI available is used to re-evaluate the postcode classification\(^5\).

In the PCI the location of a postcode unit centroid is used to link postcodes to larger geographies, such as Census Output Areas, Data Zones and Local Authorities. The PCI classification is compared to postcode classifications generated using two methods:

1) **Method 1: Location of Data Zone Centroids**
   a) Postcode units are assigned to data zones based on the location of their population weighted centroid;
   b) The location of the data zone centroid is used to link the data zone to another (higher) geography.
   c) All postcodes assigned to that data zone take the value of that data zone to assign to the higher geography.

2) **Method 2: Location of Majority of Output Areas**
   a) Postcode units are assigned to data zones based on the location of their population weighted centroid;
   b) The location of the majority of census output area centroids within a data zone is used to link the data zone to another (higher) geography.
   c) All postcodes assigned to that data zone take the value of that data zone to assign to the higher geography.

If a data zone is split in half by the boundary of a geography, the data zone can be assigned to a geography area based on the location of the population weighted data zone centroid (Method 1) or based on the location of the majority of the census output areas (Method 2). The difference between these two methods is minimal (see Table 3.1).

<table>
<thead>
<tr>
<th>Geography</th>
<th>Units</th>
<th>Method 1 Data Zone Centroids</th>
<th>Method 2 Majority Output Area</th>
</tr>
</thead>
<tbody>
<tr>
<td>Urban/Rural Classification</td>
<td>6(^5)</td>
<td>5,845</td>
<td>4.0</td>
</tr>
<tr>
<td>Health Board Areas</td>
<td>14</td>
<td>68</td>
<td>0.0</td>
</tr>
<tr>
<td>Local Authorities</td>
<td>32</td>
<td>25</td>
<td>0.0</td>
</tr>
<tr>
<td>Community Health Partnerships</td>
<td>40</td>
<td>202</td>
<td>0.1</td>
</tr>
<tr>
<td>Westminster Parliamentary Constituencies</td>
<td>59</td>
<td>2,010</td>
<td>1.4</td>
</tr>
<tr>
<td>Scottish Parliamentary Constituencies</td>
<td>73</td>
<td>2,208</td>
<td>1.5</td>
</tr>
<tr>
<td>Multi Member Wards</td>
<td>353</td>
<td>7,493</td>
<td>5.2</td>
</tr>
</tbody>
</table>

\(^5\) Note that in amendments to the evaluation, postcodes are only checked with the most recent PCI in reference to the geography that has been updated. It is felt that it is not necessary to re-run the analysis for geographies which have not changed with the most recent PCI as the processing involved is quite lengthy and the results would show little change.

\(^6\) Geographic areas of the urban rural classification are not necessarily contiguous.
In Table 3.1, method 1 and 2 are compared against the assignments in the Postcode Index. The table compares the two different look up methods which clearly show that, regardless of whether method 1 or 2 is used, the results will be very similar. This makes sense because both the output area master grid reference and the data zone centroid are population weighted, so the centroids used to assign postcodes must be similar.

In the case of the Local Authorities boundaries, which were coterminous with data zones in 2001, minor changes have been made since 2001 and currently 25 postcode units are classified inaccurately in the whole of Scotland. Based on the method described in this report this would correspond to an inaccuracy of 0.02%.

For geographies which are created with different building blocks and which comprise of relatively small area units such as Multi Member Wards, the inaccuracy caused by crossing boundaries is more substantial (5.2%, method 1), although this is a considerable improvement on the percentage inaccuracy of the previous Census based wards (14.9%).

Generally geographies that have more areas, such as Wards, seem to cause more postcodes to be classified inaccurately than geographies that have fewer areas (e.g. Health Boards). This could be explained by the fact that in the case of more areas, the geography boundaries and the data zones boundaries are likely to cross more often and be less coterminous. This is not the case, however, for the urban/rural classification on postcodes. This geography differs from the others in that it is a classification of areas rather than distinct individual units. Though there are technically 6 ‘areas’ in the 6-fold classification, these areas or not contiguous meaning that there can be a large number of polygons that could cross more frequently with data zones.

Figure 3.2 and 3.3 show the postcode units that were classified inaccurately for each of the geographies based on both method one and two look ups. Figure 3.2 gives the total number of postcodes and Figure 3.3 as a percentage of all postcodes.

![Figure 3.2: Total number of postcode units that are classified inaccurately, for seven geographies.](image)
Figure 3.3: Percent of postcode units that are classified inaccurately, for seven geographies.

In Summary:

Scottish Government Urban/Rural Classification, Scottish Government 2009/2010
This geography differs from the other six geographies, because it is a classification of areas rather than a set of individual geographic units. The 6-fold classification divides the area of Scotland into 6 categories based upon GROS Settlements and accessibility, and is updated on a biannual basis. Since Settlements are built from Postcode units, these areas are relatively coterminous with data zones, however, outside of these areas the classification is based on drive-times which do not fit well with data zones. Users of the classification are encouraged to use the census output area (and postcode) version of the classification wherever possible.

The most current version of the classification is the 2009/2010 release. The number of inaccurately classified postcodes as of the end of 2010 is 5,845 under method 1 and 5,517 using method 2 (an inaccuracy range of 3.8 – 4.0%). This is a slight increase from the last version of the classification (2007/2008), which had 5,469 to 5,640 (method 2 and 1, respectively) postcodes classed inaccurately, but a similar range of 3.8-4.0%. Map 3.1 shows the location of postcode units that currently have an inaccurate urban/rural classification when classed by data zones (method 1).

Health Board Areas, Scottish Executive 2006
This geography consists of 14 Health Board areas. The current Health Board boundaries are static and changes are bound by legislation. Health boards were based on groups of local government regions or districts in existence prior to 1996, and for the most part follow local authorities (though there are instances where the boundaries cross). Because of the large size of each Health Board in relation to a data zone, the impact of boundary changes to the accuracy of the postcode units is small.

Currently 68 postcode units, less than 0.05 % of all units, are classified inaccurately (method 1). Map 3.2 shows the location of the postcode units which currently have an inaccurate Health Board classification when classed using data zone centroids.
**Local Authorities, OS BoundaryLine 2010**
In total, Scotland is divided into 32 Local Authorities. During the creation of the data zones in 2001, the data zone boundaries were coterminous with the Local Authority boundaries. Local Authority boundaries are changed occasionally and this can affect the accuracy. Because of the large size of each authority in relation to a data zone, the crossings of data zone boundaries will be rare and no immediate problems are expected.

Currently only 25 postcode units, or less than 0.02% of all units, are classified inaccurately (method 1). Map 3.3 shows the location of the postcode units with an inaccurate Local Authority classification when classed using data zone centroids.

**Community Health Partnerships, Scottish Government 2007**
In total there are 40 Community Health Partnerships which are coterminous with Local Authority boundaries or sub-divisions of boundaries. Community Health Partnership boundaries are changed occasionally and this can affect the accuracy. Because of the large size of each Community Health Partnership in relation to a data zone, the crossings of data zone boundaries will be rare and no immediate problems are expected.

Currently between 201 and 202 postcode units (0.1%) are classified inaccurately, depending on the used look-up method. Map 3.4 shows the location of the postcode units with an inaccurate Community Health Partnership classification when classed using data zone centroids.

**Westminster Parliamentary Constituencies, OS BoundaryLine 2005**
There are 59 Westminster Parliamentary Constituencies. The boundaries of this geography are created from Electoral Wards which are non-coterminous with data zones. This geography is static and is only updated during a review, which takes place every eight to twelve years.

The number of inaccurately classified postcode units is between 1,982 and 2,010 (1.4%). Map 3.5 shows the location of the postcode units with an inaccurate Westminster Constituency classification when classed using data zone centroids.

**Scottish Parliamentary Constituencies, OS BoundaryLine 2010**
There are 73 Scottish Parliamentary Constituencies. The boundaries of this geography are created from Electoral Wards which are non-coterminous with data zones. This geography is static and is only updated during a review, which takes place every eight to twelve years. A review of the Scottish Parliamentary Constituencies was recently carried out which resulted in the release of new constituency boundaries in the later half of 2010.

The number of inaccurately classified postcode units is between 1,195 and 2,208 (0.8-1.5%). Map 3.6 shows the location of the postcode units with an inaccurate Scottish Constituency classification, when grouped using data zone centroids.

**Multi Member Wards, OS BoundaryLine 2010**
There are 353 Multi Member Wards covering Scotland. The boundaries of this geography are based on a combination of electoral population, community definitions and Ordnance Survey TOIDS. Multi Member Wards are non-coterminous with data zones. This geography is static and is only updated during a review, which takes place every eight to twelve years.

The number of inaccurately classified postcode units is between 7,307 and 7,493 (5.0-5.2%), which is a relatively bad fit. Map 3.7 shows the location of the postcode units with an inaccurate Multi Member Ward classification, based on data zone centroids.
Map 3.1: Postcodes assigned to the wrong Urban/Rural category when using data zones to group postcode units.
Map 3.2: Postcodes assigned to the wrong Health Board when using data zones to group postcode units.
Map 3.3: Postcodes assigned to the wrong Local Authority when using data zones to group postcode units.
Map 3.4: Postcodes assigned to the wrong Community Health Partnership when using data zones to group postcode units.
Map 3.5: Postcodes assigned to the wrong Westminster Parliamentary Constituency when using data zones to group postcode units.
Map 3.6: Postcodes assigned to the wrong Scottish Parliamentary Constituency when using data zones to group postcode units.
Map 3.7: Postcodes assigned to the wrong Multi Member Ward when using data zones to group postcode units.
4. Conclusions

The Scottish Government Geographic Information Science & Analysis Team has agreed to monitor issues related to changes over time on an annual basis to evaluate the viability of the static data zone boundaries. The three time-related issues are defined as demographic changes, boundary issues and loss of social homogeneity.

Demographic changes
Data zones were originally designed to have populations between 500 and 1000 people. Gradually, from 2001 to 2009, there has been a widening of the population distribution such that more data zones fall outside this optimal range. The percentage of data zones out-with the acceptable population range has grown from 6.2% in 2001 to 14.3% in 2009, based on GROS Small Area Population Estimates. The number of data zones, with a population of less than 300 or more than 1500, has increased from 12 to 89 between 2001 and 2009. In 2009, a total of 166 data zones have an estimated population of less than 500 and a total of 757 data zones have an estimated population of more than 1000.

Boundary Issues
The main cause for inaccuracy is the use of different building blocks in the definition of new geographies. If the data zone geography is to be used as a universal smallest geography unit, as currently done by SNS, then ideally, all new statistical geographies should use data zones as building blocks. An alternative could be to use postcode units, which are smaller than data zones but share similarities.

In addition, geographies that have more areas, such as Multi Member Wards seem to cause more postcode units to be classified inaccurately than geographies that have fewer areas, such as Health Board Areas. This can be explained by the more areas the geography has the more often it’s boundaries and data zones boundaries are likely to cross and be less coterminous. This of course does not apply to geographies such as the Urban/Rural Classification – which is a categorisation of areas across Scotland, rather than a set of unique units. Inaccuracies observed in the urban rural classification of postcodes by data zone centroids likely stem from the way in which the classification is created, combining settlement population data with accessibility information. Although the areas covered by settlements may fit quite well with data zones, areas which are classified by their accessibility in terms of drive-times are not coterminous with data zones at all.

Loss of Social Homogeneity
By definition, data zones were created to have uniform social composition. However, over time social and economic changes will have taken place that may impact on the homogeneity of the data zone. At this time, this issue has not yet been covered in this evaluation, however, future research will aim to assess the impacts of social changes within data zones.

---

7 Scottish Neighbourhood Statistics website: www.sns.gov.uk