2013 Consultation regarding the redraw of Data Zones

The Scottish Government
Riaghaltas na h-Alba

2013 Consultation regarding the redraw of Data Zones

Executive Summary

Data Zones are the key geography for small area statistics in Scotland and are widely used across the public and private sector. They are large enough that statistics can be presented accurately without fear of disclosure and small enough that they can be used to represent communities. Aggregations of Data Zones are often used to approximate a larger area of interest or a higher level geography that statistics wouldn't normally be available at. They have roughly standard populations to allow comparison between Data Zones and static boundaries to allow tracking of change over time.

There is however a number of issues associated with Data Zones. Since their inception in 2004, there has been population redistribution and the link between Data Zones and other geographies has become eroded. This reduces the utility of Data Zones and invalidates some of the criteria used during their initial development, hence the need to redraw Data Zones.

The 2011 Population Census will provide us with the most accurate small area statistics available for some time and provide us with the unique opportunity to redraw Data Zone boundaries.

This consultation paper sets out the background to Data Zones; why Data Zones need to be redrawn; how we have produced the draft 2011 Data Zones for consultation and details of how to see them on the Scottish Neighbourhood Statistics website; and the consultation questions

The redraw process made changes (i.e. a split, merge, or correction of a multi-extent data zone) to 706 of the original 6,505 Data Zones. The remaining 5,799 have been rebuilt using the 2011 Census Output Areas on a best-fit basis.   The total number of draft 2011 Data Zones is 6,940.

Responses to this consultation are invited before 12th February 2014.

Office of the Chief Statistician and Performance
October 2013

Scottish Government
GWR, St Andrews House
Edinburgh
EH1 3DG

Telephone: 0131 244 0442
e-mail: neighbourhood.statistics@scotland.gsi.gov.uk
internet: www.sns.gov.uk

1. Introduction

1.1. Data Zones first became available in Spring 2004 following the results of the 2001 population Census. The final publication of the Data Zone geography was seen as a milestone in the ability to monitor and develop policy at small area level.

1.2. Scotland is currently divided into 6,505 Data Zones of varying size and shape, but with a roughly standard population in both rural and urban areas. Data Zones are a static geography created purely for the purpose of presenting statistics.

1.3. Since their inception, Data Zones have become the main small area geography used for presenting statistics. They are the core geography behind the Scottish Neighbourhood Statistics website (www.sns.gov.uk) and the Scottish Index of Multiple Deprivation in addition to being the main small area geography for a wide range of statistical outputs.

1.4. Data Zones were built up from 2001 Census Output Areas (COAs) and met tight constraints on population thresholds (500 – 1,000 household residents). The aim was to build Data Zones by grouping together COAs with similar social characteristics, for Data Zones to have a fairly compact shape, and to take account of physical boundaries.

1.5. Data Zones are intended to be used as static geographic units over time. This allows analysis of change over time at a small area to be carried out, without the problem of discontinuities caused by boundary changes. There are, however, a number of issues related to temporal change that affect their validity such as:

- o Population Change: Over time, houses will be built or demolished and some Data Zones will grow or shrink beyond the original population thresholds.

- o Boundary Changes: Most geographies are not static. When other geographies change their boundaries, they often split Data Zones leading to inaccurate classifications of Census Output Areas or postcodes.

- o Loss of Social Homogeneity: The Data Zones were created to have social homogeneity. However, over time social and economic changes may have taken place that may reduce the level of social homogeneity within a Data Zone.

1.6. For the reasons noted above, Data Zones need to be redrawn. The purpose of this consultation is to gain user views on the draft set of Data Zone 2011 boundaries.

1.7. Proposed draft 2011 Data Zone boundaries can be viewed and downloaded at www.sns.gov.uk/Consultation

1.8. The closing date for providing submissions to this consultation is the 12th February 2014.

2. Background

2.1. The main criteria (constraints) used in the initial definition of Data Zones, in decreasing order of importance, were:

- approximate equality of population, between 500 and 1,000 people;
- compactness of shape;
- approximate homogeneity of social composition;
- existence, where possible of some community of interest;
- accordance with other boundaries of local significance;
- accordance with prominent features in the physical environment.

2.2. More detailed information on how Data Zones were originally created can be found at:
http://www.scotland.gov.uk/Topics/Statistics/sns/SNSRef/SNSPapDatZon

2.3. During the initial creation of the Data Zones it was acknowledged that often these constraints could not all be met. Therefore, a good compromise between these constraints was found for each Data Zone. The smallest Data Zone, located in Edinburgh (Figure 1.1, right), covered an area of only 12,367 square metres and had a total population of 501 people in 2001. The largest Data Zone was located in the Highlands and covered an area of 1,159 square kilometres with a population of 636 people in 2001.

© Crown copyright 2009. All rights reserved Scottish Government. Licence number: 100020540 2009. Scottish Government GI Science & Analysis Team 30 January 2009, Job: 4464 - KT
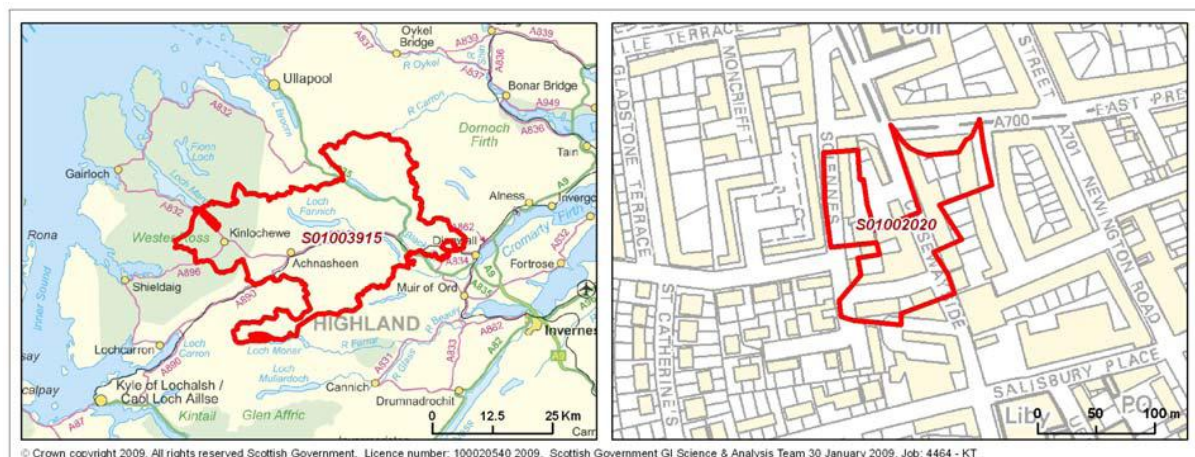
Figure 1.1: Largest (left) and smallest (right) Data Zones in terms of area extent.

2.4. In the proposed redraw, Data Zones will now be built from 2011 Census Output Areas (COAs) and will be designed so that they can be built up to form many higher level geographies using the same original constraints above. In turn, COAs are built up from postcodes (see Figure 1.2 below). Current Data Zones will remain (and will be called 2001 Data Zones) and the new Data Zones will be known as 2011 Data Zones.

Geography Hierarchy

Local Authority
↑
Intermediate Geography
↑
Data Zones
↑
Output Areas
↑
Postcodes

- Council Areas – Majority of Statistics available on LA level.
- ~4,000 residents – mainly used for data too disclosive/unreliable at data zone level.
- ~800 residents – core small area statistics geography.
- ~50 households – mainly used for the Small Area Census results.
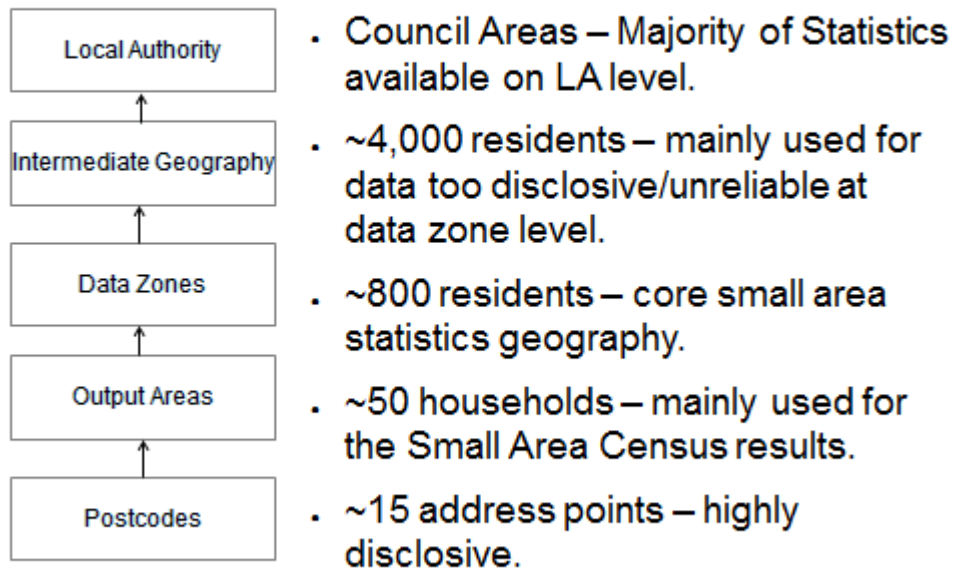- ~15 address points – highly disclosive.

Figure 1.2 Geography hierarchy.

3. The Consultation Process

3.1. The Office of the Chief Statistician and Performance are running this consultation because it is necessary to redraw Data Zones, for reasons expanded on in the following section. Draft 2011 Data Zones can be viewed and downloaded at www.sns.gov.uk/Consultation. We are seeking views on the draft Data Zones and will consider all suggestions made. We will publish the final 2011 Data Zones in mid-2014.

Why Data Zones need to be redrawn

4. Population Change

4.1. When Data Zones were first introduced, one of the main aims was to produce a static and stable geography. Maintaining the existing boundaries would be consistent with the original stated aims of Data Zones and would ensure that there are no discontinuities in any time series of statistics presented on Data Zones. The down side of this is that, where there has been significant population change, statistics presented at Data Zone level may not be as relevant as they could be.

4.2. The disadvantage of leaving Data Zone boundaries unchanged is that, where Data Zone populations have shrunk considerably, there may be disclosure issues with the statistics presented in these areas. This may mean that data might have to be withheld or rounded to prevent disclosive information from being released. In addition, where the population in an area has grown, the Data Zone may not represent a neighbourhood or small community. It also means in some cases that postcodes and Census Output Areas (COAs) need to be assigned to Data Zones on a best-fit basis, leading to inaccurately classified postcodes and COAs.

5. Changes to Higher Level Geographies

5.1. Problems can occur when the boundaries of other geographies change and common boundaries between Data Zones and higher level geographies become eroded. This means that, where in the past it would have been possible to aggregate Data Zones and get an exact match for a higher level geography, only best-fit aggregations are now possible. This also raises a potential disclosure issue where it may be possible to infer information about the very small gaps that exist between aggregations of Data Zone statistics and statistics produced using the exact boundaries of higher level geographies. These gaps are often referred to as slivers.

5.2. Local Authority boundaries rarely change, but can cause disclosure issues when they do. In figure 1.3 below, the Data Zone boundary exactly followed the Local Authority boundary in 2001. Since then, there have been some developments in the area and the Local Authority boundaries have been redrawn to accommodate this, while the Data Zone boundaries remain unchanged.

Data Zones Crossing Local Authority Boundaries

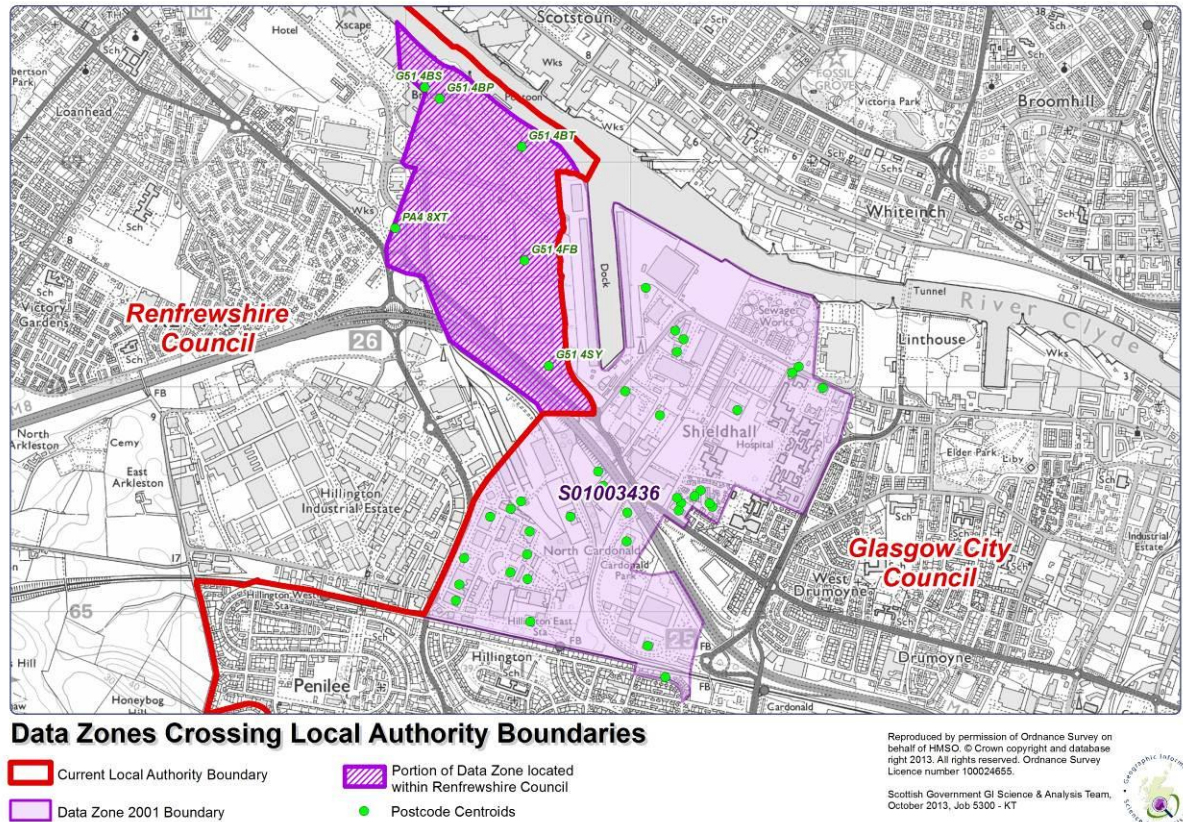|		
▭ Current Local Authority Boundary	▨ Portion of Data Zone located within Renfrewshire Council	
▭ Data Zone 2001 Boundary	● Postcode Centroids	

Fig 1.3: Showing Data Zone S01003436 which crosses the Local Authority boundary between Renfrewshire and Glasgow City

5.3. In this case, the shaded area towards the north west of the Data Zone should be allocated to Renfrewshire council, and will be if statistics are calculated on the basis of postcodes, but fall within Data Zone S01003436 and will be allocated to Glasgow City council if Data Zones are used to aggregate to Local Authority boundaries.

5.4.   If a set of statistics were available at both Data Zone level and at Local Authority level, it would be possible to deduce information about the gaps between these areas. This raises a disclosure risk and in order to prevent this, analysts need to apply disclosure control to their statistics or withhold certain figures. The type of disclosure control applied is likely to be rounding or probabilistic rounding. This creates additional work for the analyst and reduces the utility of the statistics. Further information on statistical disclosure control can be found in this paper:

http://www.scotland.gov.uk/Resource/Doc/933/0076518.doc

6. Changes to Lower Level Geographies

6.1. 2001 Census Output Areas are the building blocks of current Data Zones. In turn postcodes, as at 2001, were used as the building blocks of Census Output Areas. Data Zones will have matched 2001 postcode boundaries, but since 2001

postcode boundaries have changed, new postcodes have been brought in and old postcodes have been removed. This erodes the link between postcodes and Data Zones which means postcodes will be aggregated up to Data Zone level on a best-fit basis.

6.2. Postcodes are owned and maintained by the Royal Mail purely for the purposes of delivering mail. Analysts use postcodes as the building blocks of almost all higher level geographies including Data Zones, as it is the most common spatial referencing point available. The Royal Mail do not define postcode boundaries, however the NRS (National Records of Scotland) produce postcode boundaries based on the location of address points with the same postcode and it is these boundaries that were used during the inception of Data Zones. The NRS update the postcode boundaries every six months.

6.3.   Where postcode boundaries have changed since Data Zones were first produced, this may make Data Zone boundaries appear to have a discontinuity when new Data Zone boundaries are introduced. Although the exact boundary line will have changed, where the only change to a Data Zone is due to postcode drift, there will be no discontinuity in the majority of statistics. Since analysts will be using a postcode lookup to allocate records to Data Zones, the same records will be allocated to the Data Zone both pre and post 2011. The only effect is that the new boundaries will represent the exact area that the statistics relate to, rather than a notional area with records allocated on a best-fit basis using the location of the postcode centroid.

6.4. There are two issues associated with postcode drift that affect Data Zones. The first is that postcode boundaries will gradually change as address points are added or removed. Since analysts will usually use postcodes to allocate data to higher level geographies, this means that the notional area that the statistics relate to will gradually change over time and will be slightly different to the exact Data Zone boundaries. This affects all geographies, however it is only when you look at very small geographies that this becomes noticeable.

6.5. The other issue associated with postcode drift is related to using the postcode centroid to allocate postcodes to Data Zones. The postcode centroid is located at the address point nearest to the average easting and northing of all address points within the postcode. This, combined with the odd shape and the small size of the postcode geography, can occasionally result in postcodes switching Data Zones if there is a very small change to the address points in a postcode.

6.6. In Figure 1.4, postcode boundaries exactly matched Data Zone boundaries in 2001. By 2011, there have been a few changes to postcode boundaries and the fit has become eroded. Many of the changes are relatively small and are to be expected as address points are added or removed and notional postcode boundaries change. However, postcode DG10 9AJ has experienced some development which has resulted in a slight change to the postcode boundary. These changes have resulted in the entire postcode switching Data Zone even though only a very small part of the postcode is in a different Data Zone.
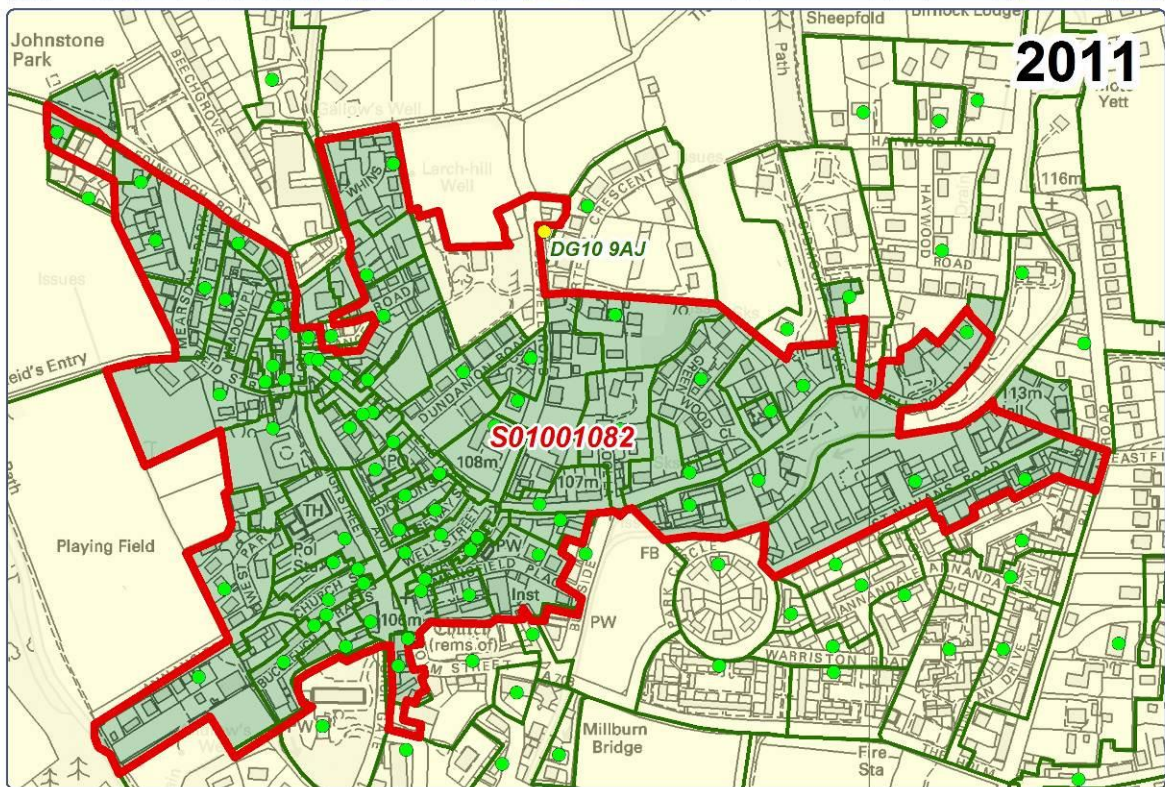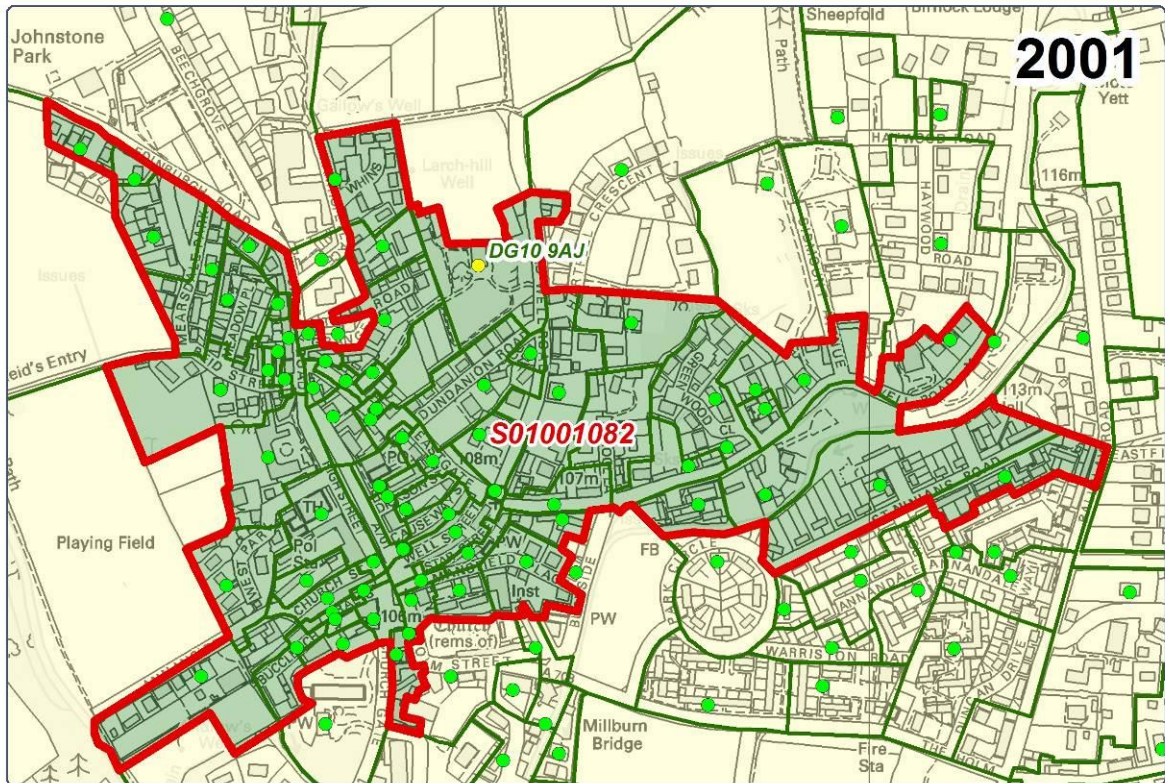
Aggregation of Postcodes to Data Zones

Scottish Government GI Science & Analysis Team
October 2013, Job 5300 - KT

Fig 1.4: Showing the effect of changes to postcode on their aggregation to Data Zone S01001082

6.7. The main implication of this is that postcodes are assigned to Data Zones on a best-fit basis and the area that the statistics relate to will be slightly different to the area within the Data Zone boundary. Although it makes sense for Data Zone boundaries to align with postcode boundaries when the Data Zone boundaries are being drawn up, this link will slowly erode over time which is why Data Zones must be redrawn.

6.8. To summarise, several factors make it beneficial to redraw the boundaries of Data Zones. These include where Data Zones have seen significant population change, where the Data Zone boundaries no longer align to key higher level geographies, where Census Output Area boundaries have changed or where there is a compelling argument to redraw the boundary to reflect local circumstances.

7. Implications of changing Data Zones to other Statistical Outputs

7.1. The two statistical outputs that will be most directly affected by changes to the Data Zone geography are the Scottish Index of Multiple Deprivation (SIMD) and the 6 and 8-fold Urban Rural Classification. These changes will result in a discontinuity for SIMD. This is unfortunate, but unavoidable, as the only way to prevent this would be to make absolutely no changes at all to Data Zones.

7.2. The Urban Rural classification is essentially a shape-file, but it is also released as a postcode lookup and a Data Zone lookup. The majority of official statistics will use the postcode lookup when producing analysis by the Urban Rural classification, and changes to Data Zones will not affect these analyses. Analysts who use the Data Zone lookup to analyse their own or other published statistics by Urban Rural classification will be affected.

7.3. Health Boards and NUTS are completely independent of Data Zones, they will not be affected by our proposals.

7.4. Data Zones have no influence over Local Authority boundaries. However Local Authority boundaries affect Data Zones.

7.5. Data Zone level statistics will not as a matter of course be produced for both existing Data Zones and 2011 Data Zones. Previous statistics will not be recalculated for 2011 Data Zones, as a matter of course.

7.6. Table 1.1 describes the level of change by Local Authority. The highest level of change was seen in Glasgow, where 15.4% of all data zones that changed were located. South Lanarkshire, Aberdeenshire, Fife, Edinburgh, and North Lanarkshire each account for between 5 and 10% of the total change, While Orkney, Shetland, East Dunbartonshire, East Renfrewshire, and North Ayrshire each accounted for less than 1% of the total change. No changes were required in Eilean Siar after the initial best-fit aggregation of 2011 COAs.

7.7. Of the 976 Data Zones identified as 'most deprived' on the Scottish Index of Multiple Deprivation (SIMD, 2012), 113 of these have changed. The greatest

proportion of these (47) are located in Glasgow, representing 41.6% of total changes to most deprived data zones.

Table 1.1: 2001 Data Zone Changes

Council Area	Total 2001 No. Data Zones	Total No. Changed	% of Change w/in Council	% of Total Change
Aberdeen City	267	25	9.4%	3.5%
Aberdeenshire	301	42	14.0%	5.9%
Angus	142	15	10.6%	2.1%
Argyll & Bute	122	10	8.2%	1.4%
Clackmannanshire	64	11	17.2%	1.6%
Dumfries & Galloway	193	16	8.3%	2.3%
Dundee City	179	17	9.5%	2.4%
East Ayrshire	154	23	14.9%	3.3%
East Dunbartonshire	127	6	4.7%	0.8%
East Lothian	120	11	9.2%	1.6%
East Renfrewshire	120	7	5.8%	1.0%
Edinburgh, City of	549	50	9.1%	7.1%
Eilean Siar	36	0	0.0%	0.0%
Falkirk	197	22	11.2%	3.1%
Fife	453	50	11.0%	7.1%
Glasgow City	694	109	15.7%	15.4%
Highland	292	18	6.2%	2.5%
Inverclyde	110	12	10.9%	1.7%
Midlothian	112	17	15.2%	2.4%
Moray	116	13	11.2%	1.8%
North Ayrshire	179	7	3.9%	1.0%
North Lanarkshire	418	64	15.3%	9.1%
Orkney Islands	27	1	3.7%	0.1%
Perth & Kinross	175	25	14.3%	3.5%
Renfrewshire	214	20	9.3%	2.8%
Scottish Borders	130	13	10.0%	1.8%
Shetland Islands	30	2	6.7%	0.3%
South Ayrshire	147	9	6.1%	1.3%
South Lanarkshire	398	40	10.1%	5.7%
Stirling	110	9	8.2%	1.3%
West Dunbartonshire	118	11	9.3%	1.6%
West Lothian	211	31	14.7%	4.4%
Scotland	**6505**	**706**	**10.9%**	100.0%

How we have produced draft 2011 Data Zones

8. Census Output Areas

8.1. In 2001, Census Output Areas (COAs) were used as the building blocks to produce Data Zone boundaries.

8.2. As COAs are used as building blocks for Data Zones, any changes to COAs will have knock on effects for Data Zones.

8.3. In the redraw process, 2011 Census Output Areas (COAs) were aggregated up to existing Data Zone boundaries on a best-fit basis to produce a first cut of updated Data Zones. Where the population of the Data Zone is outwith the accepted population range, that Data Zone has then been split into 2 or more Data Zones or merged with a neighbouring Data Zone in order to maintain roughly standard populations. These splits were made along Census Output Area boundaries and, where possible, any merges have been made within Intermediate Zone boundaries.

8.4. The new Data Zones will be called 2011 Data Zones, while the existing Data Zones would then be referred to as 2001 Data Zones.

8.5 Postcodes are allocated to Data Zones based on the location of the postcode centroid. This means that postcodes are essentially allocated to Data Zones on a best-fit basis. 2011 Census Output Areas respect postcode boundaries, so aligning Data Zone boundaries to 2011 Census Output Areas has the added benefit that Data Zone boundaries will also align to 2011 postcode boundaries.

9. Changes to Census Output Areas

9.1. The process of creating boundaries for output areas for the 2011 Census was largely automated. Each postcode in the underlying postcode geography at the point when it is frozen for the purposes of the 2011 Census enumeration (January 2011) where allocated initially to its 2001 Census Output Area. Some COAs have been merged with a neighbouring COA (e.g. where housing demolition means their population has dipped below confidentiality thresholds, causing a disclosure risk), and some have been split (e.g. where new housing has taken them above maximum population thresholds.) This approach is designed to ensure as much comparability as possible between the 2001 and 2011 output geographies.

9.2. 2011 postcodes have been allocated to 2001 COA boundaries on a best-fit basis. These aggregations have formed the first cut of 2011 COAs. The 2011 COAs were then aligned to Local Authority (LA) boundaries, as some postcodes cross LA boundaries and some LA boundaries have changed since 2001. COAs outwith the existing population ranges have been split into 2 or more COAs or merged with a neighbouring COA to ensure a roughly standard population across all COAs. Where a split or merge has occurred, it has been made along a locally

recognised boundary. A hierarchy of boundaries has been used to determine where the new boundaries should fall where a COA is being split or merged. As Data Zones are built up from 2001 COAs, 2011 COAs should have a good level of comparability with 2001 Data Zones.

10. Methodology for producing draft 2011 Data Zones

10.1. Following a consultation in August 2010, the decision was made that Data Zones would be redrawn using 2011 Census Output Areas. The aim of the redraw process was to maintain the link with current Census data, to re-align to Local Authority boundaries, and to account for changes in population while keeping boundaries relatively comparable to those created in 2001. The main criteria (constraints) used to redraw Data Zones for 2011, in approximate order of importance (descending), were:

- That they be built up from 2011 Census Output Areas;
- Maintain approximately equal populations of 500 to 1,000 people, with an absolute minimum of 375 people and a maximum population of 1,125;
- Be a continuous area without multiple extents, unless it contains islands;
- Where possible, changes are made within Intermediate Zone boundaries;
- Maintain a compactness of shape.

10.2. The production of the draft 2011 Data Zones was largely an automated process. (If you require a detailed methodology, please contact us.)

10.3. At the outset, a first cut of boundaries was created by aggregating 2011 Census Output Areas to 2001 Data Zones, based on the location of the COAs population weighted centroid. The resident population was summed for each 'best-fit' Data Zone to assess whether it fell within the required population thresholds.

10.4. Data Zones that have seen a dramatic decrease in population (e.g. as the result of a demolition) were merged with a neighbouring Data Zone (this effected 23 Data Zones). Conversely, Data Zones that have grown significantly in population (e.g. as the result of a new housing development in the area) were split into one or more new Data Zones, this has effected 364 Data Zones. The population thresholds to which either a merge or split was required are 375 and 1,125 respectively.

10.5. Where a merge was required, a Data Zone with a population of less than 375 was merged to a neighbouring Data Zone with the longest shared boundary, maintaining the population criteria where possible. This effected 23 data zones.

10.6. For Data Zones that required splitting (364); Census Output Areas were grouped into 'sub' Data Zones that meet the population requirements. Starting with the most south westerly, Census Output Areas were appended on to create a new Data Zone until the resident population totalled 750 or more. At the point, a new 'sub' Data Zone was created, and the remaining Census Output Areas

appended on until the population threshold was reached. The number of new Data Zones that were created from a split Data Zone was calculated by dividing the resident population by 750 and rounding to the nearest whole number.

2001 Data Zones to be 'Split' or 'Merged'
Data Zones that have a resident population of over 1,125 will be split into multiple data zones, while those with a resident population below 375 will be merged with a neighbouring data zone in the redraw process.
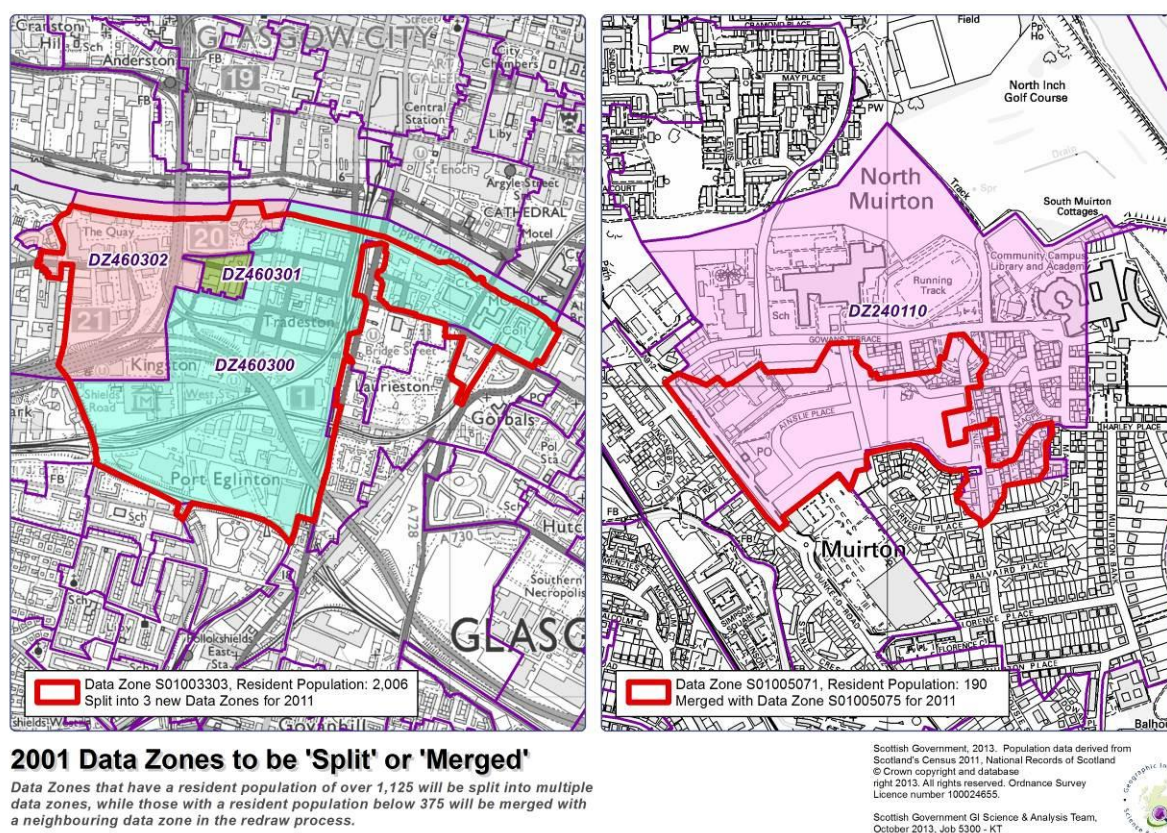
Fig 1.5: Examples of Data Zones that have been split or merged

10.7. Figure 1.5 shows examples of Data Zones that have been split or merged. Building up from 2011 Census Output Areas, Data Zone S01003303 (left) has a resident population of 2,006 - well above the desired population range of 500-1,000. In the draft set of 2011 Data Zones produced, this area has been split into 3 new Data Zones (labelled DZ460300, DZ460301, and DZ460302).

10.8. Data Zone S01005071 (Figure 1.5, right) currently has a resident population of 190, which is well below the minimum desired population. To correct this in the draft set of 2011 Data Zones, S01005071 has been merged with S01005075 (labelled DZ240110).

10.9. In some cases, where the Census Output Area geography has significantly changed, multi-extent Data Zones were created after aggregating Census Output Areas to 2001 Data Zones. Where this occurred, some additional editing was required to correct this. An example of this is shown in Figure 1.6, where 2001 Data Zone S01000163 is cut across by the 2011 Census Output Area, creating two parts of the Data Zone that do not touch. In this case, Census Output Area S00090409 was removed from Data Zone S01000163 and appended to 2011 Data Zone DZ330151. Changes made as a result of multi-extents affected 319 Data Zones.

Multi-extent Data Zones

Multi-extent data zones can be created when aggregating 2011 Census Output Areas to 2001 Data Zones using the location of their centroid, as the data zone may be cut across by an output area not linked to that data zone.

Data Zone S01000163 | 2011 OAs assigned to S01000163 by best-fit | OA cutting across, but not assigned to S01000163 | OA Centroid
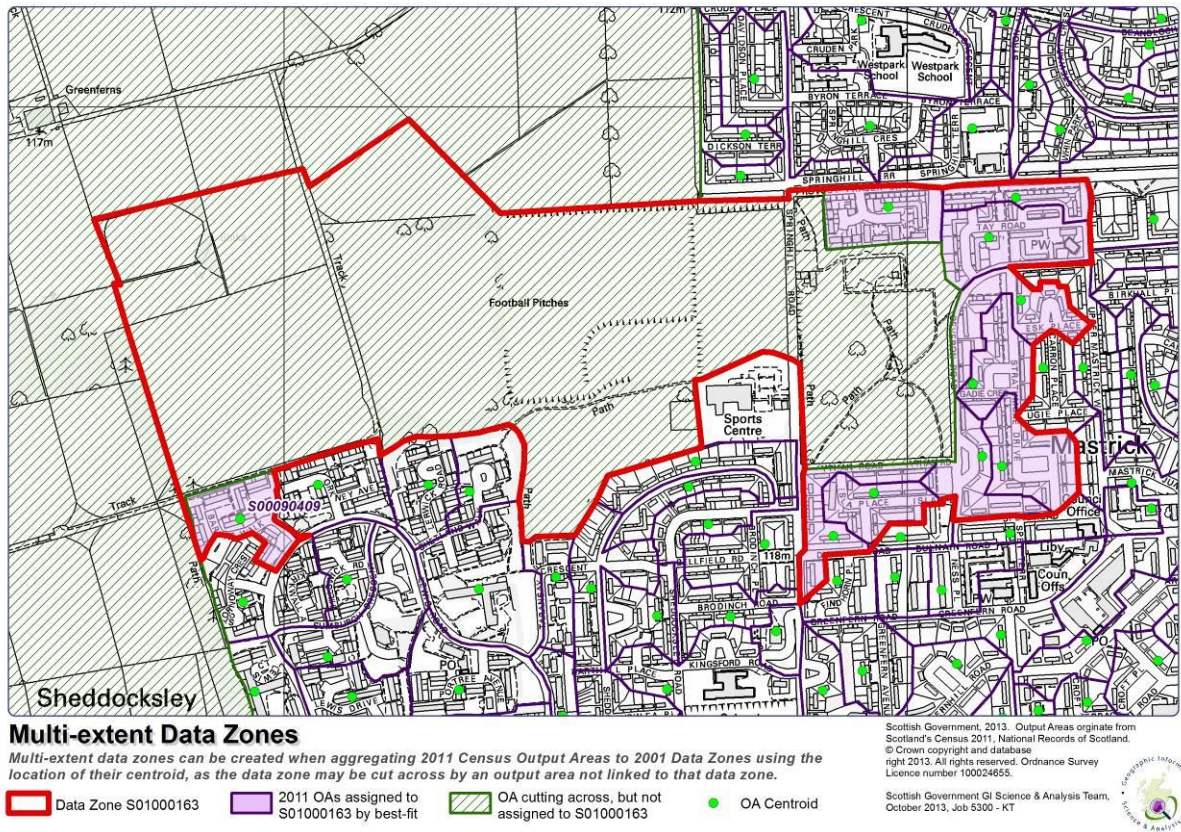
Fig 1.6: Example of a Multi-extent Data Zone created by building 2011 Census Output Areas to 2001 Data Zones on a best-fit basis

10.10. The redraw process made changes (i.e. a split, merge, or correction of a multi-extent data zone) to 706 of the original 6,505 Data Zones. The remaining 5,799 have been rebuilt using the 2011 Census Output Areas on a best-fit basis. The total number of 2011 Data Zones is 6,940.

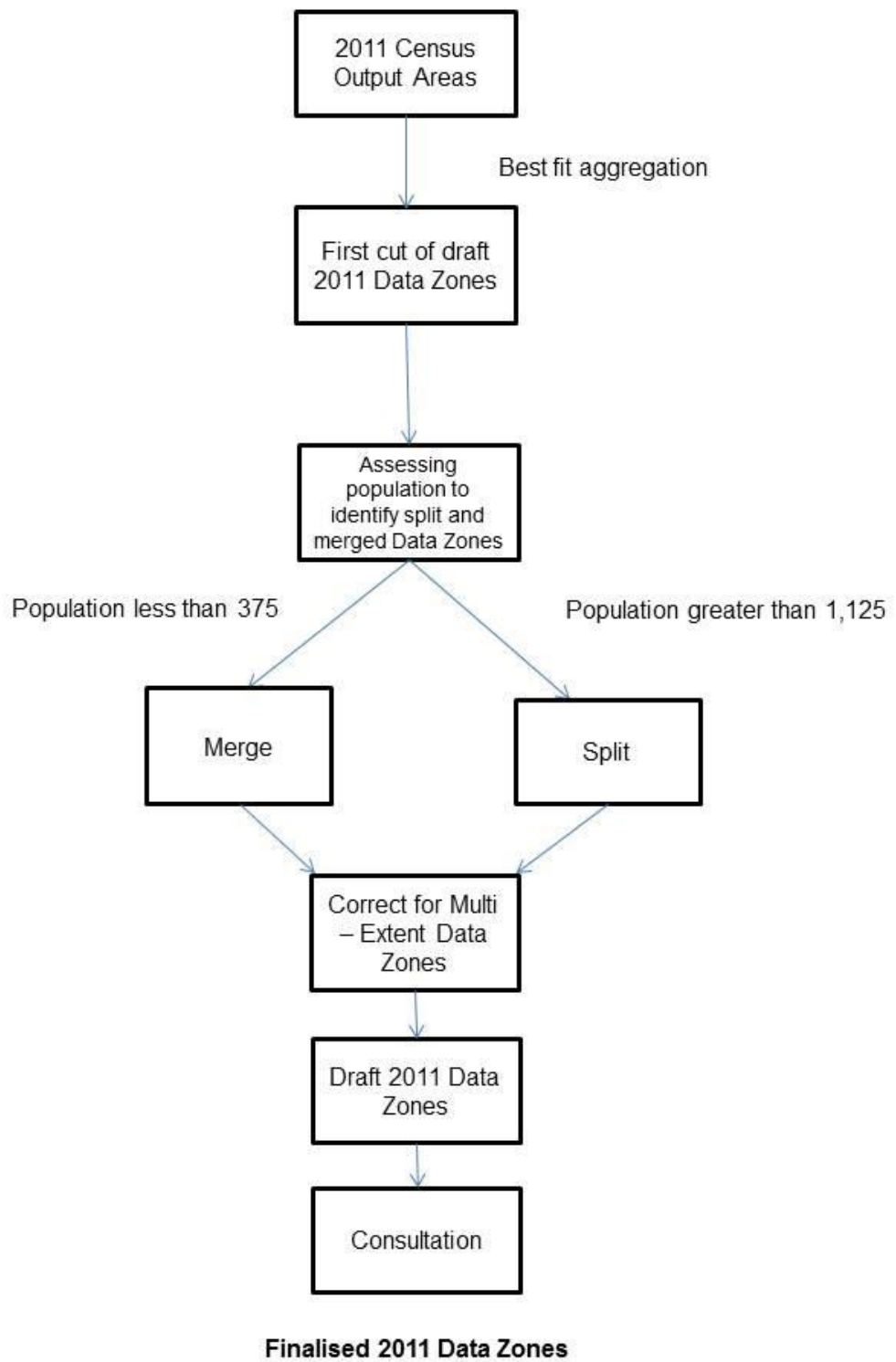10.11. Figure 1.7 shows a high level summary of methodology described above.

Finalised 2011 Data Zones

Figure 1.7: High level methodology for draft 2011 Data Zones

11. Calculation of the Data Zone Centroid

11.1.   The Data Zone centroid represents the centre of the area. This is not the geometric middle of the Data Zone, but a point that represents the population centre. The main use of the centroid is to determine which higher level geography the Data Zone would be allocated to. Figure 1.8 shows an example of this, where a 2011 Data Zone (labelled DZ360288) crosses Multi Member Ward boundaries of City Centre, Southside/Newington, and Meadows/Morningside in Edinburgh. In this case, the Data Zone centroid falls within the Southside/Newington ward, so for reporting and aggregation purposes would be assigned to this ward.
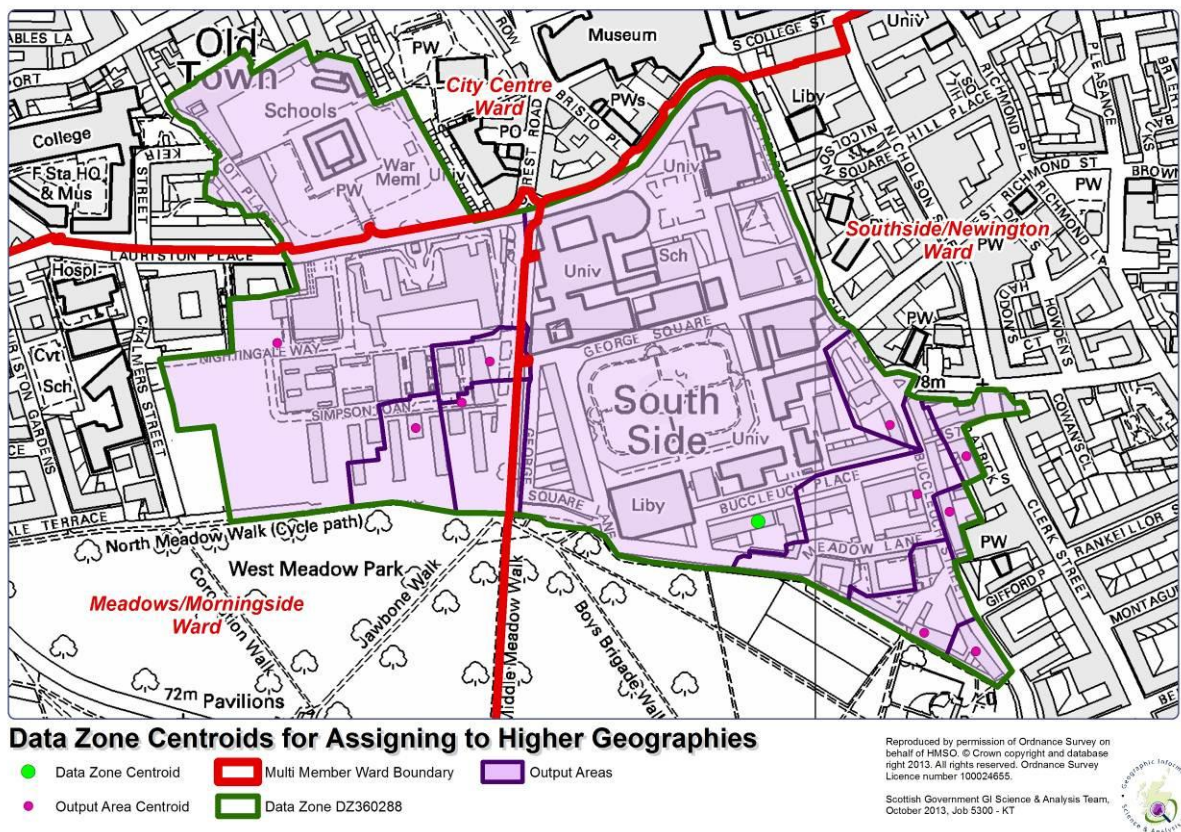

Fig 1.8: The centroid is used to assign Data Zones to higher geographies

11.2. For 2001 Data Zones, centroids were calculated as the population weighted centre (essentially the mean centre) of all 2001 Census Output Areas contained within the Data Zone. The methodology used can be found here:

http://www.scotland.gov.uk/Resource/Doc/933/0082884.doc

11.3. For 2011 Data Zones, we are proposing a small change in the methodology used to calculate the Data Zone centroid. The methodology remains broadly the same, but instead of using the population weighted mean, we are proposing to use the output area median.

11.4. The median is a measure of central tendency and, broadly speaking, the median can be thought of as the 'middle' value. While the mean is calculated by

17

summing all the values together and then dividing by the number of observations, the median is calculated by putting the observations in order, from lowest to highest, and then taking the value in the middle. (Or calculating the mean of the two middle values if there are an even number of observations.).

11.5. The key advantage of using the median is that it is not as heavily influenced by extreme values as the mean. If a Data Zone has a highly skewed population distribution, for example a large rural data zone containing a small town in one corner, then the mean can be heavily influenced by the small number of people who live far away from the population centre and the mean will likely fall outside of the town. The median is considered to be a more robust measure of central tendency and is less likely to be influenced by values far away from what would be considered to be the population centre of the Data Zone.

11.6. The process for creating 2011 Data Zone Centroids was automated using ESRI ArcGIS, but the general method is described below:

1)   The median easting and northing coordinate pair for all Census Output Areas within the Data Zone is calculated, giving a notional centroid of the Data Zone.

2)   Data zones can be complex shapes, so to ensure that the median falls within the data zone boundary, a second step is carried out move it to the nearest output area centroid. The distance from each of the Census Output Area centroid to the median easting/northing is calculated using Pythagoras' Theorem. The Census Output Area coordinate pair with the shortest distance to the median was chosen to represent the centroid of the Data Zone.

11.7. For the purposes of the consultation, centroids using both the mean and median methods have been created for review. Of the 6,940 centroids, 1,431 differ in location between the mean and median methods. Differences occur in all councils except the Western Isles and Glasgow. The council with the greatest discrepancy is Dumfries and Galloway, where 35% of centroids fall in a different location depending on whether the mean or median method is used.

11.8. The differences in the locations of the centroid may have an impact on the Data Zones assignment to a higher level geography. For example, 6 of the 2011 Data Zones would be assigned to a different 2001 Intermediate Zone, and 28 would be assigned to a different Multi Member Ward.

11.9. Figure 1.9 shows two examples of where the mean and median centroid could impact higher geography assignments. The 2011 Data Zone on the left hand map (labelled DZ340098) would be assigned to the Westhill North and South Intermediate Zone using the mean centroid, however, it would be assigned to Garlogie and Elrick if the median centroid is used. The map on the right shows how data zone DZ260047 would be assigned to the Jedburgh and District Multi Member ward if the mean centroid is used, but would be assigned to Kelso and District ward if the median is used.
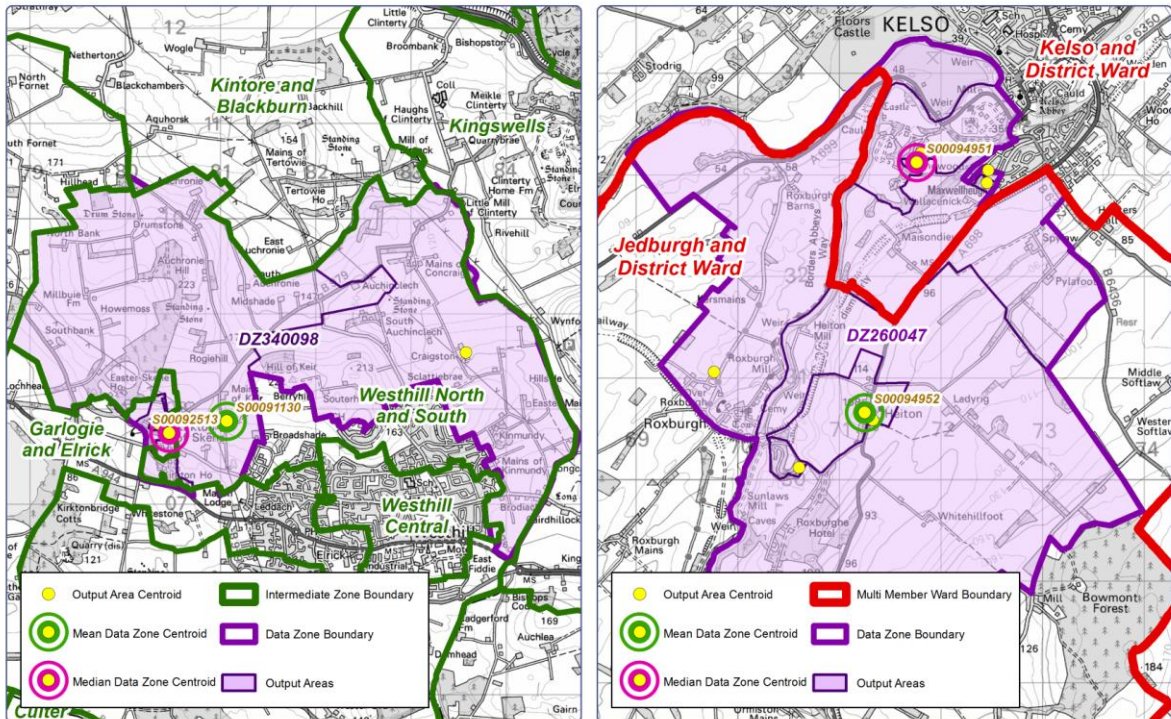
Mean and Median Data Zone Centroids

Using the mean or median method to calculate the Data Zone centroid can have an impact on its assignment to a higher geography. The examples above show where a Data Zone could be assigned to a different Intermediate Zone or Multi Member Ward depending on which centroid method is used.

Scottish Government, 2013. © Crown copyright and database right 2013. All rights reserved. Ordnance Survey Licence number 100024655.

Scottish Government GI Science & Analysis Team, October 2013, Job 5300 - KT

Fig 1.9: The mean and median centroids may lead to differences in higher geography assignments for some data zones

11.10. Consultation Question:

 Do you agree that 2011 Data Zones should use the median methodology?

Other aspects for consideration

12. Intermediate Zones

12.1. Intermediate Zones (sometimes called Intermediate Geographies) are similar to Data Zones and share many of the same traits and problems. The key difference is their size: Intermediate Zones have a population of around 4,000 and are built up from Data Zones. Intermediate Zones tend to be used when the statistics are too sensitive at Data Zone level. The larger populations of Intermediate Zones mean that statistics published on them are less likely to require disclosure control.

12.2. Intermediate Zones will continue to be aggregates of Data Zones and the approach taken with Data Zones will determine the approach to be taken with Intermediate Zones. The population ranges that Intermediate Zones should conform to will remain between 2,500 and 6,000.

12.3. Due to Data Zones being proposed in draft format, best fit Intermediate Zones have been created. Intermediate Zones will be named and finalised after this consultation. Best fit Intermediate Zones can be viewed along with the proposed Data Zones through the consultation web mapping service.

12.4. Consultation Question:

Are you content with the proposed best fit Intermediate Zones?
If changes occur to the proposed 2011 Data Zones post consultation, these changes will be reflected in the Intermediate Zones.

13. Data Zone Codes

13.1. 2011 Data Zones will continue to use the S01 standard code, but all Data Zones will be given new codes starting at S01006506. Users should note that the S0 codes will not be assigned until after the consultation is complete to allow for possible changes to draft Data Zones. In the meantime, temporary codes have been assigned using the format 'DZ000000'. The codes are grouped together so that all Data Zones within a Local Authority will have consecutive codes. A look-up table is provided using the temporary codes so that users can map 2001 Data Zones to 2011 Data Zones.

13.2. The current coding system for Data Zones is a 9-character code of the format ANNNNNNNN (A=letter, N=number), where the first three characters represent the geography type (here S01 indicates Data Zone).

 13.3. After the consultation is complete Data Zones without names will be named after the Intermediate Zone that they fall within, using the naming convention IZ name – 01, IZ name – 02, etc. A look-up table is provided so that users can map 2001 Data Zones names to draft 2011 Data Zones. (Data Zones

previously named will have names transferred to 2011 Data Zones once finalised.)

13.4. Data Zone codes will be allocated after this consultation once Data Zone boundaries have been finalised.

14. Overview

- 2011 Data Zones have been built up from 2011 Census Output Areas, which in turn have been built up from 2011 Postcodes;

- Data Zones are aligned to Local Authority boundaries;

- Where the population of a Data Zone has grown or shrunk significantly, it has been split into two or more areas, or merged with a neighbouring Data Zone;

- Where a muti-extent Data Zone has been created in the best-fit aggregation of Census Output Areas, the Data Zone has been modified to correct for this (unless it is due to the inclusion of islands);

- Where possible, merges have occurred within Intermediate Zones' boundaries;

The reasons for taking this approach are:

- Exact 2011 Population Census results are available on the 2011 Data Zones

- The methodology is consistent with that used when Data Zones were first created;

- Data Zones will once again have roughly standard populations;

- Aggregations of Data Zones can be used to produce a comparable area over time;

- Minimal changes will be made to Intermediate Zones.

15. Future Reviews of Data Zones

15.1. The proposal is to continue to review Data Zones once every ten years.

16. Consultation

16.1. The purpose of this consultation is to receive views on the proposed redraw of Data Zone boundaries.

16.2. Consultation questions can be found at the end of this document. Respondents are encouraged to respond using the response form in Annex 1.

16.3.   Proposed draft 2011 Data Zone boundaries with mean and median centroids and best fit Intermediate Zones can be viewed at www.sns.gov.uk/Consultation. Data can be downloaded from here also.

Responding to this Consultation

The closing date for providing submissions to this consultation is the 12th February 2014. Should you wish to discuss any queries in relation to this consultation documentation please do not hesitate to contact us:

by email: neighbourhood.statistics@scotland.gsi.gov.uk

in writing: Consultation regarding the redraw of Data Zone,
                   Office of the Chief Statistician and Performance
                   GWR
                   St. Andrews House
                   Regent Road
                   Edinburgh
                   EH1 3DG

by phone: 0131 244 0442

A copy of the response form can be found below.

Consultation regarding the redraw of Data Zones

RESPONDENT INFORMATION FORM
Please Note this form **must** be returned with your response to ensure that we handle your response appropriately

1. Name/Organisation
Organisation Name

Title Mr ☐ Ms ☐ Mrs ☐ Miss ☐ Dr ☐   Please tick as appropriate

Surname

Forename

2. Postal Address

Postcode	**Phone**	**Email**

3. Permissions - I am responding as…

Individual	/	**Group/Organisation**
☐	Please tick as appropriate	☐

(a)	Do you agree to your response being made available to the public (in Scottish Government library and/or on the Scottish Government web site)?	**(c)**	The name and address of your organisation **will be** made available to the public (in the Scottish Government library and/or on the Scottish Government web site).
	Please tick as appropriate ☐ Yes ☐ No		
(b)	Where confidentiality is not requested, we will make your responses available to the public on the following basis		Are you content for your **response** to be made available?
	Please tick ONE of the following boxes		**Please tick as appropriate** ☐ Yes ☐ No

Yes, make my response, name and address all available ☐

or

Yes, make my response available, but not my name and address ☐

or

Yes, make my response and name available, but not my address ☐

(d) We will share your response internally with other Scottish Government policy teams who may be addressing the issues you discuss. They may wish to contact you again in the future, but we require your permission to do so. Are you content for Scottish Government to contact you again in relation to this consultation exercise?

Please tick as appropriate ☐ **Yes** ☐**No**

CONSULTATION QUESTIONS

Are you content with the proposed 2011 Data Zones?

Yes ☐ No ☐

If you wish to make suggestions for change to a small number of draft Data Zones please provide the Data Zone code together with an explanation of its design limitations and the statistical benefits that would result from altering its boundary. Any accompanying maps and future supporting details would also be useful; preferably, a list of Census Output Areas with their current draft Data Zone assignment, along with the proposed new Data Zone assignment should be included.

Comments

Do you agree that 2011 Data Zones should use the median methodology for the calculation of centroids?

Please see page 17 for further information.

Yes ☐ No ☐

Comments

Are you content with the proposed best fit 2011 Intermediate Zones?
If changes occur to the proposed 2011 Data Zones post consultation, these changes will be reflected in the Intermediate Zones.

Yes ☐ No ☐

If you wish to make suggestions for change to a small number of best fit Intermediate Zones; please provide the Intermediate Zone code together with an explanation of its design limitations and the statistical benefits that would result from altering its boundary. Any accompanying maps and future supporting details would also be useful; preferably, a list of Draft 2011 Data Zones with their current draft 2001

Intermediate Zone assignment, along with the proposed new Intermediate Zone assignment should be included.

```
┌─────────────────────────────────────────────────────────────┐
│                                                             │
│                                                             │
│   Comments                                                  │
│                                                             │
│                                                             │
└─────────────────────────────────────────────────────────────┘
```

Annex 1.

List of Consultees:

Responses to this consultation are welcome from any individual or organisation with an interest. The following have been invited to respond

Convention of Scottish Local Authorities;
Local Authority Chief Executives;
Scottish Members of European Parliament;
Equalities and Human Rights Commission;
Scottish Enterprise;
NHS Health Boards;
Scottish Fire and Rescue Service;
Police Scotland

w w w . s c o t l a n d . g o v . u k