

# Scottish Neighbourhood Statistics Data Zones

## Background Information

### Introduction

Over the last couple of years the Scottish Executive have consulted widely with users on the development of a small area geography (data zones) for use in Scottish Neighbourhood Statistics to allow statistics across a number of policy areas to be readily (and regularly) available on a consistent and stable geography.

The publication of the data zones is a significant milestone in the Scottish Executive and partners' ability to monitor and develop policy at a small area level. Through Scottish Neighbourhood Statistics data zones will increasingly be the core geography for making available small area statistics across most policy areas including information about benefits, education, health and the labour market. This will allow users to readily bring together information from various sources on a common geography. Statistics relating to data zones will become available through Scottish Neighbourhood Statistics during Spring 2004.

The data zone geography covers the whole of Scotland and nests within local authority boundaries. Data zones are groups of Census output areas which have populations of between 500 and 1,000 household residents, and some effort has been made to respect physical boundaries. In addition, they have compact shape and contain households with similar social characteristics.

This report sets out background information on the development and use of data zones and also aims to answer the questions which arose during the final phase of consultation with local authorities.

Office of the Chief Statistician  
February 2004

Scottish Executive  
3WR, St Andrews House  
Edinburgh  
EH1 3DG

Telephone: 0131 244 0442  
Fax: 0131 244 0335  
e-mail: [statistics.enquiries@scotland.gsi.gov.uk](mailto:statistics.enquiries@scotland.gsi.gov.uk)  
internet: [www.sns.gov.uk](http://www.sns.gov.uk)

<b>Contents</b>	<b>Page</b>
<b>Introduction</b>	
<b>1 Frequently Asked Questions</b>	<b>1</b>
<b>2 Boundary Guidance</b>	<b>4</b>
<b>3 Data zone Index Guidance</b>	<b>6</b>
<b>ANNEX A: Methodology – Report from St Andrews University</b>	<b>8</b>
<b>ANNEX B: Scottish Neighbourhood Statistics - Data Zones Licence</b>	<b>41</b>

## **1 Frequently Asked Questions**

### **How will the Scottish Executive use data zones?**

Data zones will play a key role in Scottish Neighbourhood Statistics and the Scottish Multiple Index of Deprivation to help develop and monitor Scottish Executive and its partners' policies, particularly in terms of area regeneration. Data zones will increasingly be **the** core geography for making available small area statistics across most social topics including information about benefits, education, health and the labour market. For the first time, this will allow users to readily (and regularly) bring together information from various sources on a common small area geography.

### **Do data zones represent communities on the ground?**

It was always recognised that data zones are statistical areas and do not necessarily delineate communities. The key feature of data zones is that they are significantly smaller than previous geographies for which statistics have been available (postcode sector or ward) and are much more effective in identifying small areas with particular social characteristics, and are also more flexible in aggregating to specific areas of user interest. While at the same time data zones are large enough to protect confidentiality and to allow regular updates to be made available.

### **How were data zones produced?**

Data zones meet tight constraints on population thresholds (500 – 1000 household residents), they all nest in to local authorities and are built up from 2001 Census output areas. The aim was also to build data zones by grouping together output areas with similar social characteristics, for data zones to have a fairly compact shape, and to take account of physical boundaries. To do this (in most areas) we used primary school catchment areas to get a first cut of data zones and refined these using a measure of local homogeneity (based on employment levels, over-crowding and car-ownership) and underlying geographical features. A report from St Andrews describing the process for producing data zones is included at Annex A.

### **How were users consulted?**

Over the last couple of years the Scottish Executive consulted widely on the development of data zones. A Scottish Neighbourhood Statistics Neighbourhood Definitions Group was established in 2001 to oversee the development of data zones which included representatives from local authorities. The project was split into five phases.

The first phase invited interested parties to develop strategies for the production of data zones and this was followed by a review of the various strategies by St Andrews University. The next two stages involved St Andrews producing a prototype set of data zones for the whole of Scotland based on 1991 Census information and users throughout Scotland commenting on the prototype set. The prototype set of data zones were generally well received and a review of the first four phases was presented at a seminar in Glasgow at the end of 2002.

The main findings from the first four phases were that the methodology devised by St Andrews could be used to produce data zones and that a final set of data zones should be created using 2001 Census information. After the 2001 Census information became available in 2003, St Andrews produced a final set of data zones and invited local authorities to comment on the data zones and provide suggested amendments based on local knowledge. At this stage we were expecting minimal change, however a small number of local authorities suggested significant changes to the data zones in their areas.

**Were all suggested changes to data zones accepted?**

All suggested changes that were consistent with the tight constraints on population were accepted. Some local authorities suggested changes to data zones so that they better respected settlement boundaries or some administrative geographies.

**What effect have the accepted changes based on local knowledge had?**

Because there were a number of criteria to be met, there was no one optimal solution and there would always be some trade-offs between the various criteria. However, it is clear that constraining the data zone populations tightly has had the largest effect on the overall set of data zones. Any changes resulting from local knowledge is likely to have resulted in only a very marginal improvement in the set of data zones in terms of local homogeneity and/or respecting various boundaries.

**Was enough time made available for the final phase of consultation?**

St Andrews provided the data zones to local authorities as they became available and in the final phase local authorities were given 4 to 6 weeks to comment on the data zones. This final phase of consultation coincided with holiday period and some local authorities asked for the consultation to be extended. The consultation was further extended from mid October to mid-November.

**Which boundaries do data zones respect?**

Data zones have been designed to respect local authority boundaries as at 2001 Census.

During the consultation some users asked for data zones to take account of ward, old district, or settlement boundaries. There was no Scotland wide consensus as to which boundaries to take account of. A practical Scotland wide solution was needed and increasing the number of boundaries to be respected would have made it increasingly difficult to produce data zones (increasing the number of different boundaries to be respected would have an impact on the meeting other constraints).

During the final phase of consultation some local authorities suggested specific amendments to data zones so that they respected better some of these listed boundaries. These changes were accepted where they met the tight constraints, however, future changes to those listed boundaries will not in itself direct change to data zone boundaries.

**Are 'doughnut' shaped data zones acceptable?**

We have tried to avoid the creation of 'doughnut' shaped data zones. However, adjustments to data zones which result in doughnut data zones have been accepted where they are based on local knowledge.

**Will data zones be a stable geography?**

The intention is for data zones to be a stable geography. The Scottish Executive are currently developing the process for maintaining data zones. With each yearly update of Scottish Neighbourhood Statistics we will need to ensure that data zones meet the minimum population thresholds. Once data zones have bedded in to the Scottish Neighbourhood Statistics we intend to undertake an evaluation which will feed into the longer term process for maintaining data zones.

**How will data zones take account of population change?**

As part of Scottish Neighbourhood Statistics the General Register for Scotland (GROS) is developing a methodology to produce population estimates at the data zone level. 2001 Census population estimates were used in the creation of data zones.

**Will data zones take account of current and future planning applications?**

Information on population change as a result of development and regeneration will come from the General Register for Scotland small area population estimates.

**Will care be needed when analysing trends in data zone that have had relatively large changes in population?**

Yes. The Scottish Executive are considering developing an indicator to flag up data zones which have had significant population change.

**Will data zones be named?**

There are 6,505 data zones in Scotland and it is not practicable to name each of the zones. Each data zone will be assigned an individual code; and each code will have attributes including relationships to some other higher area geographies. Data zones are constructed from census output areas (whose boundaries are fixed) and the historic link between census output areas and data zones will be maintained.

**Why did you use 2001 Census output area as the geography to build data zones?**

There are various operational and practical reasons why the 2001 Census output area geography was used to build the data zones. These include Census output areas are readily available and cover the whole of Scotland; Census output area boundaries are fixed; a great deal of information is now postcode referenced and the links between postcodes and census outputs areas are maintained. This allows ready aggregation of postcoded information to data zones via the Census output area.

**Will the links between postcodes and data zones be maintained?**

The links between postcodes and data zones are needed to allow statistics to be produced for SNS. The intention is to continue to make the links available to SNS users. GROS will also maintain the relationships on the GROS Postcode Index.

**What data will be made available at the data zone level?**

Given confidentiality constraints are overcome, the intention is to provide data across all Scottish Neighbourhood Statistics topics at the data zone level. See [www.sns.gov](http://www.sns.gov) for the current topics included. The Scottish Executive have recognised that not all data can be made available at the data zone level and have plans to develop an intermediate geography based on aggregations of data zones during 2004.

**Are there any confidentiality issues with using data zones?**

The Scottish Executive statisticians work to a Code of Practice which prevents us from releasing analyses which could identify specific individuals. We use several methods to prevent this, including rounding the results.

**What other geographies will be supported in Scottish Neighbourhood Statistics?**

The key geography for disseminating results will be the data zone, results will also be available at the Local Authority, Health Board and Scottish Parliamentary Constituency level. During 2004 the Scottish Executive will undertake a consultation on an intermediate geography that will be aggregated from data zones.

**How do I get a copy of the data zone boundaries and look-up tables?**

The CD containing the information can be obtained by emailing your contact and address details to [statistics.enquiries@scotland.gsi.gov.uk](mailto:statistics.enquiries@scotland.gsi.gov.uk)

## 2 Boundary Guidance

<b>Title</b>	Scottish Neighbourhood Statistics Data zones
<b>Abstract</b>	Data zones is the core geography for dissemination of results in Scottish Neighbourhood Statistics (SNS). Data zones are based on 2001 Census nesting within Council areas, built up from complete 2001 output areas and have population between 500 and 1000. Data zones are a stable geography and can be used to analyse changes over time.
<b>Subject</b>	Administrative
<b>Keywords</b>	Society
<b>Data Rights</b>	Unrestricted
<b>Supplier Name</b>	Scottish Executive
<b>Supplier Contact Name</b>	Geographical Information Service Data Manager
<b>Supplier Address Line 1</b>	1 - J88
<b>Supplier Address Line 2</b>	Victoria Quay
<b>Supplier Address Line 3</b>	Edinburgh
<b>Supplier Country</b>	Scotland
<b>Supplier Postcode</b>	EH6 6QQ
<b>Supplier Telephone</b>	0131 244 1441
<b>Supplier Fax</b>	0131 244 1443
<b>Supplier Email</b>	jane.sanderson@scotland.gsi.gov.uk
<b>Supplier Web Address</b>	www.scotland.gov.uk
<b>Date Created</b>	FEB 2004
<b>Update Frequency</b>	Irregular
<b>Format</b>	Shape File
<b>Size (Mb)</b>	15
<b>Language</b>	English (UK)
<b>Location</b>	SEGIS
<b>Source</b>	GROS 2001 Census Output Area
<b>Project</b>	Scottish Neighbourhood Statistics
<b>Type</b>	GIS vector polygons
<b>Geographical Extent</b>	Scotland
<b>Special Reference System</b>	British National Grid
<b>Lineage</b>	Primary School catchments approximated to output area boundaries which are then amalgamated to data zones based on population, compactness and

	social homogeneity
<b>Positional Accuracy</b>	In line with GROS Census 2001 Output Areas
<b>Attribute Accuracy</b>	Good
<b>Completeness</b>	Complete
<b>Creator Name</b>	Scottish Executive
<b>Creator Contact Name</b>	Geographical Information Service Data Manager
<b>Creator Address Line 1</b>	Area 1-J88
<b>Creator Address Line 2</b>	Victoria Quay
<b>Creator Address Line 3</b>	Edinburgh
<b>Country</b>	Scotland
<b>Creator Postcode</b>	EH6 6QQ
<b>Creator Telephone</b>	0131 244 1441
<b>Creator Fax</b>	0131 244 1443
<b>Creator Email</b>	jane.sanderson@scotland.gsi.gov.uk
<b>Creator Web Address</b>	www.scotland.gov.uk

### 3 Data zone Index Guidance

The Data zone Index is made up of four components:

- SNS Postcode Index – the postcode look-up file which shows the geographic relationships between postcodes and census output areas, data zones and local authorities.
- SNS Census Output Area Index – the look-up showing the relationships between 2001 census output areas and data zones and local authorities.
- SNS Data zone Index – the look-up showing the relationships between data zones and local authorities.
- Names and codes – standard names and codes for geographies supported in Scottish Neighbourhood Statistics.

#### Data sources

The following data sources have been used to produce the data zone index

GROS Postcode Index 2003v2  
GROS 2001 Census OA INDEX  
SNS Data zones

#### (A) SNS Postcode Index (snspsc03bp)

The SNS Postcode Index is the postcode look-up file which shows the geographic relationships between postcodes and some of the geographies supported in Scottish Neighbourhood Statistics. The SNS Postcode Index differs from the publicly available GROS postcode index because within SNS relationships are determined via 2001 Census output areas. This means that the postcode relationships are inherited from the census output area rather than being determined by identifying which area the actual postcode unit falls in.

The SNS Postcode Index is created by combining (i) the postcode unit to census output area relationships from the 2003v2 GROS Postcode Index, with (ii) the 2001 Census output area to SNS Geographies from the 2001 Census OA Index; and (iii) SNS data zones.

#### *Fields on the index*

The index includes the following relationships and fields:

Postcode unit (outward code)	(postcod1)
Postcode unit (inward code)	(postcod2)
Postcode unit to 2001 Census output area	(outcd01)
Postcode unit to data zone	(data_zone)
Postcode unit to local authority	(council_area)
Date of introduction	(dateintr)
Date of deletion	(datedele)
Large user identifier	(splitare)

### *Current and deleted postcodes*

The SNS Postcode Index includes both current and deleted postcodes.

### *Large user postcodes*

Large user postcodes are those postcodes which receive more than around 50 mail items per day. The SNS Postcode Index includes large user postcodes and these are identified through the large user identifier field (splitare). The following codes are used in splitare:

- A=split postcode
- Z=large postcode matched with one of the split areas for a small postcode;
- L =large postcode matched with small postcode - no split areas;

### *Split postcodes*

A very small number of postcodes straddle local authority boundaries. When this happens GROS (who maintain boundaries for postcodes) split the postcode in to two or more parts and label these A, B, C etc. Split postcodes have been stripped out of the SNS Postcode Index and only the largest part (in terms of households) is included in the index.

## **(B) SNS Census Output Area Index (snscoa03bp)**

SNS Census Output Area Index – the look-up showing the relationships between 2001 census output areas and some of the SNS geographies.

### *Fields on the index*

Census output area	(outcd01)
Census output area to data zone	(data_zone)
Census output area to local authority	(council_area)

## **(C) SNS Data zone Index (snsdz03bp)**

The SNS Data zone Index provides some key information about data zones including their relationships to local authorities.

Data zone	(data_zone)
Data zone area measurement (hectare)	(dzarea)
Data zone to local authority	(council_area)

Data zones have been coded sequentially from 1 through 6505, using the following structure:

S01000001

...

S01006505

S – represents Scotland

01 – represents the geography i.e. 'data zone'

001234 – represents the data zone number

## **(D) SNS Names and codes**

SNS Names and codes sets out the names and codes to be used for local authorities, census area statistic wards, Scottish Parliamentary Constituencies and health boards within SNS.

**ANNEX A: Methodology – Report from St Andrews University**



**Report to the Scottish Executive**

**THE PRODUCTION OF  
AN UPDATED SET OF DATA ZONES  
TO INCORPORATE 2001 CENSUS  
GEOGRAPHY AND DATA**

**Professor Robin Flowerdew  
Dr Elspeth Graham  
Dr Zhiqiang Feng**

**School of Geography and Geosciences  
University of St Andrews**

**27 January 2004**

# **The production of an updated set of data zones to incorporate 2001 census geography and data**

This report describes work by Professor Robin Flowerdew, Dr Elspeth Graham, Dr Zhiqiang Feng and David Manley of the School of Geography and Geosciences, University of St Andrews, St Andrews, Fife KY16 9AL. Robin Flowerdew is Professor of Human Geography, Dr Graham is Reader in Geography, Dr Feng is Research Fellow, and Mr Manley is a postgraduate student. Communications about the report should be directed to Professor Flowerdew (telephone 01334 463853; FAX 01334 463949; email: [rf15@st-andrews.ac.uk](mailto:rf15@st-andrews.ac.uk)).

## **Contents**

<b>1</b>	<b>The project</b>	<b>3</b>
<b>2</b>	<b>Background</b>	<b>4</b>
<b>3</b>	<b>Methods</b>	<b>5</b>
	<b>Design of data zones</b>	<b>5</b>
	<b>Consultation with local authorities</b>	<b>7</b>
	<b>Quality assurance</b>	<b>8</b>
	<b>Output</b>	<b>8</b>
<b>4</b>	<b>Results</b>	<b>9</b>
<b>5</b>	<b>Results from the consultation exercise</b>	<b>11</b>
<b>6</b>	<b>Conclusion</b>	<b>17</b>
	<b>Figures</b>	<b>18</b>
	<b>Appendix 1</b>	<b>27</b>
	<b>Appendix 2</b>	<b>29</b>
	<b>References</b>	<b>32</b>

## **The project**

The Scottish Neighbourhood Statistics project is intended to provide small-area data on a range of topics, including population, social conditions, housing, health, crime and education. Geographical units must be defined for which the data can be provided, ideally at a micro-level so users can aggregate the units to approximate places that they have an interest in. Suitable units must therefore be defined; these are referred to as data zones, rather than neighbourhoods, because the latter term suggests some degree of community identification with the zones, which is difficult or impossible to achieve in a consistent manner across Scotland. Data zones should be composed of contiguous whole census output areas, and should fit within local authority boundaries. The following criteria were taken into account in the definition of data zones, in approximate order of importance:

- 1) approximate equality of population, between 500 and 1000 people;
- 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; and
- 6) accordance with prominent features in the physical environment.

In practice, it was impossible to satisfy all these criteria at once, and compromise was necessary.

In Phase 2 of the Neighbourhood Definition Project in 2002, provisional data zones were produced. However, in some cases there were problems arising from population and social change since the 1991 census, from which the data were taken. With 2001 census data and geographical units becoming available, it was both necessary and important to revise these zones.

Local knowledge is still important in evaluating the data zones created, and it is highly appropriate that people with extensive local knowledge should examine the zones and make suggestions for their improvement. For this purpose, maps of proposed data zones were made available to local authority staff for their comments and suggestions.

It is also the case that the data zones created as part of this project may require modification at a later stage, as a result of further residential development or demolition or local government boundary changes. The Scottish Executive will be responsible for updating the data zone system and clear and complete guidance information will be given to help it to do so.

## Background

For the last few decades, the availability of accurate and systematic small-area social, economic and demographic data has been effectively confined to the Small Area Statistics (and Local Base Statistics) produced for the decennial population census. The census is restricted in the topics that it covers, and takes a long time from data collection to publication, as a result being effectively anywhere from two to twelve years out of date. An increasing amount of data is now stored on an ongoing basis by central and local government, in many cases with postcodes or other geographical referencing systems attached. Such data can be made available at the local level, subject to suitable constraints to preserve privacy, and the Neighbourhood Statistics programme, in England and Wales and in Scotland, is intended to create a system to allow easy access to these data for the public. However, there is not one consistent small area geography across government social statistics, and neither wards nor postcode sectors are fully satisfactory.

For the reasons of privacy mentioned above, it is not practical to make the data available for very small geographical units. On the other hand, to allow users to approximate geographical areas they are interested in, it is useful to have small units for which data can be aggregated flexibly. Defining these areas is, however, a substantial problem. In Scotland, it was the subject of a Neighbourhood Definition Project, funded by the Scottish Executive in 2001-02.

The first phase of the project took the form of funding five local authorities (or consortia of local authorities) to develop their own system for designing data zones for neighbourhood statistics. Phase 2, for which St Andrews won the contract, involved evaluating the five sets of suggestions and developing a system incorporating the best of all five. Such a system was developed, based in part on the recommendations of the Scottish Borders and Fife groups. Data zones were designated for all local authorities in Scotland. They were sent to the local authorities for their comments and suggestions, both on the method used and on the results in their own areas.

The main problem identified in this exercise was that the data zones were based on 1991 census data, and in some cases they now have populations far too big to be comparable with the rest of the data zones; furthermore some of them had very diverse social composition. It became obvious that geographical change and development since 1991 had to be taken into account. In addition, it seemed to make little sense to build the data zones on the basis of 1991 census geography; although some efforts have been made in designing Output Areas, where practicable, to preserve the same units for 2001 as were used in 1991. There were also a number of minor improvements that could be made to the methodology, such as the use of household population rather than total population, and the use of a modified version of the Townsend deprivation index to assess localised social homogeneity.

## Methods

### *Design of data zones*

The key element in the methodology developed for creating the data zones was the use of a geographical information system as a decision support tool. Most of the criteria used in defining the zones can be shown in map form, enabling the operator to display overlays of the relevant information on a screen simultaneously. This allowed informed decisions to be made immediately. It was also possible to make provisional decisions about the construction of data zones, investigate their wider implications for neighbouring areas, and reverse the initial decisions if necessary.

The objective of the methodology was to combine 2001 census output areas (OAs) into larger units which are between 500 and 1000 in population (excluding people not living in households). These data zones must themselves nest within local authority areas. They must also, as far as can be managed, meet the criteria of compactness, homogeneity and accordance with other boundaries and environmental features.

There are many ways of constructing data zones of the required population size (500-1000 people living in households). As a starting point, we used the boundaries of non-denominational primary school catchment areas, following the approach of the Scottish Borders consortium adopted in the previous data zone definition exercise. Denominational primary school catchments were not considered. In some local authorities, catchment area boundaries were not available in digitised form. In these cases, Voronoi (Thiessen) polygons were constructed around the primary schools to approximate the catchment areas. This method enabled every location in the local authority to be assigned to the geographically nearest primary school. Other complications arose where catchment areas were shared or undefined in parts of the local authority or where they straddle local authority boundaries. However, use of actual or approximate primary school catchment areas was possible for almost all local authorities. It would have been possible to use ward, settlement or other boundaries as a starting point, but there was no consensus, and a method was needed that could be used all over Scotland.

The first stage of the data zone definition process was to approximate the boundaries of each public primary school catchment area with OA boundaries. This was done by assigning each OA to the catchment area in which its centroid is located. The approximated catchment area boundaries were then considered to assess whether the resulting collection of OAs was too small, too large, or of the right magnitude to be a data zone. Where they were too small, the catchment area was amalgamated with one or more of its neighbours until the minimum size was reached. Where they were too large, the catchment area was split into sections on the basis of compactness, homogeneity and accordance with other boundaries and environmental features. If the resulting sections form appropriate zones in terms of target population and the other features, they could become data zones. However, they could also be amalgamated with parts or the whole of a neighbouring catchment area if doing so produced a more satisfactory set of data zones for the area as a whole.

Non-denominational primary school catchment areas were therefore used as a starting point in the data zone definition procedure, but their boundaries were not necessarily

sacrosanct. Usually data zones lie within one primary school catchment area, but there are many cases where they cross catchment area boundaries, if this results in a better solution in terms of the stated criteria.

This procedure required operational definitions of several terms. The population equality criterion (between 500 and 1000 people in households) was regarded as very important, and all data zones are in this population range except where there are special circumstances (see the Results section below).

Compactness of shape has several aspects. Multi-extent areas are avoided, that is data zones cannot consist of two non-contiguous sets of OAs, with the exception of islands where this is inevitable. Cases where one data zone was entirely surrounded by another were also avoided; such configurations are referred to as 'doughnuts'. However, as noted below, some doughnuts were created at the final stage as a result of user consultation. Other aspects of compactness can be measured in a number of ways, but for the purpose of this exercise the measure used was the perimeter squared divided by the area. In addition the compactness of data zones was assessed subjectively by the operator. Because the shapes of OAs are sometimes far from compact, data zones could not be as compact as might be ideal, but it was considered that a degree of compactness was important if only to help people work out which houses, streets and villages are within a data zone.

Homogeneity is a complex concept, which could be defined in terms of a wide variety of social, economic, demographic or environmental variables; for the purpose of this project, consideration was restricted to socio-economic homogeneity. This was assessed using the widely known Townsend index of deprivation (Townsend *et al.* 1988), used as a local measure of homogeneity to group together data zones with similar social characteristics. The index is defined as the sum of four standardised percentage variables from the census – unemployment, households without cars, households who are not owner-occupiers, and households in overcrowded conditions (more than 1 person per room). The index was constructed for each OA. The homogeneity of a data zone was measured by seeing how similar the Townsend scores were for each OA in the data zone (more details are given below). A minor change to the methodology used in the earlier project was that students were excluded from the unemployment calculation.

Boundaries considered in visual inspection of the draft data zones, in addition to non-denominational primary school catchment areas, included the boundaries of wards, community councils, postcode sectors and settlements. Environmental features considered include lakes and rivers, major topographic features, main roads, railways and industrial sites. Data zones do not always respect these boundaries or environmental features if doing so resulted in infringements of the criteria discussed earlier. Community attitudes and allegiances were considered only through the local authority input to the data zone design process.

The final step in the procedure was to allocate unique identifiers to the data zones produced. It is impractical to develop a set of names for zones with a household population as small as 500-1000 in urban areas, and hence codes were used. These codes were assigned according to rules established in association with the Scottish Office. Each data zone has a unique numerical code; the codes are arranged by local

authority, the local authorities being in alphabetical order. Within local authorities, the zones are ordered from south to north, according to the grid reference of the data zone centroid. A look-up table is provided listing the OAs associated with each data zone. Where there are cases of OAs consisting of two or more polygons as represented within the GIS, whole OAs are allocated to data zones, even if this means ignoring the contiguity constraint.

Most of this procedure has been automated using an amalgamation program written in FORTRAN by Zhiqiang Feng. The steps in this procedure are summarised in Figure 1. It should be stressed, however, that each of the data zones produced by the program was then examined carefully by the operator. If this examination revealed problems in the data zones produced in terms of any of the criteria, the operator could consider the potential for overcoming the problem by moving OAs between data zones. Here it was sometimes necessary to accept a trade-off between different criteria – for example, to decide whether social homogeneity was more or less important than compactness of shape.

It should be clear from the account of the methodology above that personal judgement was an important aspect of the procedure. Progress has been made on systems of zone design that are entirely computer-generated, such as the ZoDE system under development at the University of Newcastle upon Tyne (Alvanides *et al.* 2002) and the system developed at the University of Southampton (Martin 2002). However, it was considered that these were not yet suitable for the current purpose. This was either because the objective function maximised by the software was not sufficiently flexible, or because there were restrictive limits on the number of input zones that can be handled. It may be that one of these systems will be sufficiently reliable and sufficiently thoroughly tested to be used in a few years' time, but at present it is felt that there are advantages to be gained from the 'hands-on' approach we propose.

#### *Consultation with Local Authorities*

The design of data zones proceeded by considering a small number of local authorities at a time. When data zones had been developed for a particular local authority, they were sent to the local authority concerned for comments and suggestions. They were supplied in the form of a shape file that can be read by most GIS systems, together with a look-up table relating the data zones to OAs within the authority. We went through a similar procedure with the sets of zones defined in the trial exercise in 2002, receiving many helpful comments from the local authorities.

Local authorities were originally given approximately 4 weeks to comment, the precise dates depending on when the provisional data zones were sent out. However, the deadline for receiving comments was extended to 14 November 2003 and in practice comments were received and taken note of throughout November and into the start of December. Comments concerning recent changes (since the 2001 census) in particular areas were used to update the data zones. Comments about projected developments were noted for subsequent data zone revisions but it was decided that they should not be used in data zone definition before the development takes place. Comments arguing for specific changes in the data zone system for other reasons (such as accordance with other systems developed for parts of the local authority) were considered; appropriate adjustments were made unless suggested changes led to

violation of the principles established for the exercise as a whole. Such adjustments were discussed with the Scottish Executive as appropriate. Comments arguing for a different approach to the definition of data zones were noted, but were not used to adjust the data zone system.

### *Quality assurance*

This was an important element of the exercise. People responsible for designing the data zone system were in possession of at least Masters-level qualifications in GIS and were fully briefed in the methods to be employed. They were able to consult with experienced workers on issues and problems that arose during the course of the work.

Following completion of the set of data zones for a local authority, the resulting map and data base were checked by an experienced worker. These checks included ensuring that each OA within the local authority had been assigned to one and only one data zone and that each data zone had a household population between 500 and 1000. In addition, each data zone on the map was examined to check for compactness and homogeneity. If there appeared to be improvements possible to the data zone system, the checker consulted with the person who produced the original zones to determine what the best solution would be. Cases of uncertainty were referred to the project manager. The quality assurance process continued while local authorities were examining the proposed data zones, concentrating on any issues or difficulties identified by the local authorities.

### *Output*

The output from the project consists of the report, the data zones themselves, and a look-up table, which gives the identifier and composition of each data zone in terms of the OAs which make it up. Those data zones containing communal establishments are identified in the table, together with the household population, the communal establishment population and the total population. Boundaries were supplied to the Scottish Executive as shape files in ESRI software.

## Results

The eventual set of data zones was produced according to the methods outlined above. Before consultation, there were 6576 data zones. The consultation process eventually resulted in a set of 6505 data zones.

Following discussion with the Scottish Executive, it was agreed that under special circumstances it was acceptable to create zones with household populations as low as 475 and as high as 1100. This was in situations where the resulting data zone made more sense than aggregating simply to meet the threshold. As a result, 36 data zones have household populations below 500, the smallest having 476 people. There are 213 data zones with household populations over 1000, the largest having 1099 people. Thus 3.82% of the data zones are outside the original 500 – 1000 household population range.

The total populations of the data zones (including persons in communal establishments) are sometimes considerably in excess of the household populations. There are 357 data zones with total populations over 1000, and 58 over 1100; 3 of these have over 2000 total population, all of which have large numbers in communal establishments. Two of these are in Edinburgh City and one in Argyll and Bute. Only 25 data zones have under 500 people in total.

Figure 3 shows the number of data zones in each Scottish local authority, together with the average household population and the average total population. The number of zones ranges from 27 in Orkney and 30 in Shetland to 694 in Glasgow City and 549 in Edinburgh City. Given the initial target zone size of between 500 and 1000 population in households, it is not surprising that average household population is close to 750 for most authorities; indeed the median value is 750.89. There is a tendency for larger authorities to have slightly larger data zones – the average household population for Edinburgh is 797.44 and for Glasgow 815.52. However the largest average belongs to East Dunbartonshire (842.24). Highland (703.26) and Orkney (704.96) have the smallest average household population.

The use of the 2140 non-denominational primary school catchment areas as a starting point for the data zone construction process has affected but not determined the outcome. This can be seen in Figure 4, which shows that about 77% of the data zones are entirely within one primary school catchment area, while 23% overlap primary school catchment area boundaries. This analysis includes authorities where catchment areas were approximated by Voronoi polygons, but excludes the three island authorities of Eilean Siar, Orkney and Shetland, where Voronoi polygons were less appropriate because straight-line distances could be poor indicators of accessibility within and between islands.

Social homogeneity was assessed using the Townsend index modified to exclude student unemployment. It is difficult to assess the homogeneity of the data zones, but an attempt was made to do so using the following methodology. Each of the 42,604 Output Areas in Scotland was assigned a Townsend score. These scores were then divided into deciles (ten groups with equal numbers of OAs in each group). Data zones on average contain about 6.5 Output Areas. For each data zone, it is possible to

find the OA with the highest score and the OA with the lowest score. A homogeneity score is then calculated by subtracting the number of the decile containing the lowest score from the number of the decile containing the highest score. This can range from 0, if all OAs in a data zone are in the same Townsend decile, to 9 if a data zone contains an OA in the top decile and another OA in the bottom decile.

Figure 5a shows the difference in deciles that occur for all the data zones. Low values indicate a greater degree of social homogeneity. In evaluating the results, however, it should be noted that a random assignment of decile numbers to sets of six OAs would result in a preponderance of differences at the top end of the scale – more than half would be expected to have a difference of 7 or greater. Likewise, Figure 5b shows the results of the same calculations for electoral wards in Scotland. It can be seen that the data zones are considerably more homogeneous than wards.

The observed pattern therefore indicates considerable success in achieving a high degree of social homogeneity for the data zones. Experience suggests that a greater degree of homogeneity would be difficult to achieve, purely because of the geography of deprivation in Scotland. In some places, there are small areas of high relative deprivation which have to be merged with more prosperous areas in order to construct data zones of appropriate size. Because the most affluent and most deprived (even more so) are not always all grouped together in large areas, it is impossible to define completely homogeneous areas. Data zones with a higher degree of internal homogeneity could have been achieved if the population target was smaller, but the size of data zones needs to be balanced against the availability of (non-Census) data and the usability of the data zones. Our interpretation is that, considering all constraints, we have a set of data zones with as high a degree of social homogeneity as could reasonably be expected.

Figure 6 illustrates the extent to which we were successful in creating compact zones, and how compactness varies between local authorities. The concept of shape is difficult to capture successfully by a single measure, but it is common to use the ratio of perimeter squared to area as a way of measuring at least one aspect of the concept. Low values represent more compact shapes. The table shows that typical values of this statistic are in the 40s. The mean for the whole of Scotland is 43.82, with a standard deviation of 16.77; the range is from 15.00 to 188.13. For comparison, a perfect circle would have a value of about 12.56. The histogram shows a highly skewed distribution, with most data zones having relatively low values and a small number having very high values.

The table shows that some local authorities have very high average values, headed by Shetland at 77.34 and Eilean Siar at 58.14, followed by Orkney and Argyll & Bute. The highest values are generally for coastal or island areas, where the perimeter is lengthened by elongated lochs and peninsulas. Low values usually occur for smaller and more densely populated authorities that are inland or have a relatively straight coastline. The lowest average values are attained in Midlothian (37.08) and West Lothian (37.22), followed by Renfrewshire and Dundee City. Very compact data zones are in some cases difficult to construct, perhaps because the geography of social deprivation does not have a very compact pattern, or simply because the Output Areas, from which data zones are built up, may have highly elongated or convoluted shapes.

## **Results from the consultation exercise**

Data zones were produced for each of the 32 local authorities in Scotland. They were sent to the Chief Executive of each authority with a covering letter from Robert Williams (Scottish Executive) and Robin Flowerdew (University of St Andrews). This letter (see Appendix 1) explained the project and asked for feedback about the zones, including both general comments about the method used and specific comments about the zones suggested. A further letter (see Appendix 2) was sent out to local authorities by Peter Whitehouse (Scottish Executive) clarifying the exercise further and extending the initial deadline for comments. Local authorities that had not responded by November were sent a reminder by electronic mail.

Responses were received from all authorities except five (Clackmannanshire, East Lothian, Moray, Orkney and West Lothian). Many of the responding authorities made suggestions about specific changes to data zones within their boundaries, and several suggested major changes, a few supplying their own suggested data zones. These were evaluated sympathetically, and suggested changes were incorporated when they did not infringe any of the principles used in defining the data zones.

The following paragraphs summarise the main points of responses from each authority.

### **Aberdeen City**

Generally supportive of the work; a number of changes were suggested in order to bring the data zones into closer accord with the system of neighbourhood units used for planning purposes. These changes were accepted, subject to the population size constraints.

### **Aberdeenshire**

Critical of the exercise, particularly on the basis that the data zones did not accord with statistical units currently in use in Aberdeenshire. Particular criticisms of the draft data zones supplied included data zones straddling primary school catchment boundaries, straddling ward boundaries, splitting settlements and having odd shapes. Illustrative examples of these problems were given. Within our overall guidelines, we suggested alterations to meet some of these criticisms.

### **Angus**

Adjustments to the draft data zones were suggested, based on local knowledge, intended to increase coincidence with the small area boundaries used in Angus. The suggestions were adopted.

### **Argyll & Bute**

Comments were made in relation to the rural and island communities in this area. The suggestion that data zones with household populations slightly below 500 should be allowed was adopted. The suggestion that data zones could be made more acceptable by moving Output Area boundaries in unpopulated zones, however, was rejected.

because it was in conflict with our instruction that data zones must be made up of complete Output Areas. Specific suggestions for data zone revisions were made and accepted.

### **Dumfries & Galloway**

Services are organised according to the boundaries of the former district councils and the area committees. A request was made that data zones should nest within these units, and revisions were made to meet this request.

### **Dundee City**

Several suggestions for changes were made, based mainly on two criteria. Requests to change the zonal system where new developments were planned were not acceded to, on the grounds that the system had to be based purely on 2001 census household populations. However, in a few places it was possible to make minor changes so that zones expected to grow in the near future were adjusted to have relatively small populations, though still within the 500-1000 target. Requests for changes in order to increase social homogeneity resulted in a number of zone boundary changes, without compromising size and other constraints.

### **East Ayrshire**

There was concern about rural / urban definitions, in particular where rural and urban areas were grouped together in the same zone. This problem, which was raised by several other authorities too, arose from the use of primary school catchment areas as a starting point – the catchment areas usually including surrounding rural areas as well as the settlement where the school was located. Sometimes a settlement may include two schools, whose catchment areas split both the settlement and the rural area in half. Some authorities stated a preference for a system that unified the settlement and separated it from the rural area. Such a preference sometimes results in data zones shaped like doughnuts, with a more urban zone entirely surrounded by a rural zone. Our initial decision was not to accept doughnut-shaped data zones on grounds of shape, but we eventually decided to allow such zones where the local authority felt they would be useful. We also re-examined the zones in and around all settlements of 3000 population or more, giving more emphasis to settlement boundaries in the zone design process. This was also done in other local authorities where similar concerns were expressed.

### **East Dunbartonshire**

Several specific problems with the data zones were mentioned and boundaries were revised to reduce these problems. This council expressed the view that it would have liked to redraw the boundaries more extensively had more time been available.

### **East Renfrewshire**

The council was 'reasonably pleased' with the suggested data zones but suggested three minor changes, which were duly made.

## **Edinburgh, City of**

The suggested data zones were well received, although concern was expressed that there was not a closer fit to school catchment areas.

## **Eilean Siar**

A specific suggestion for boundary changes was made and acted upon. More general concerns about the treatment of islands were raised, arguing that islands with populations below the 500 constraint should be treated as independent data zones. As indicated above, it was decided to allow islands with household populations slightly below 500 to be independent data zones.

## **Falkirk**

This council was highly critical of the provisional data zones sent to them, on the grounds that they did not respect local communities or major environmental features. Council officials felt that it was necessary to redefine the data zones themselves. The resulting zones met the main criteria used in the study and were accepted

## **Fife**

A generally positive response to the exercise – ‘a valuable and enjoyable project which I am sure will take neighbourhood analysis forward in leaps and bounds’. Some changes were suggested in two areas, which were duly incorporated.

## **Glasgow City**

No specific alterations were proposed, but the view was expressed that zones of the size suggested were too small and too numerous to be appropriate for Glasgow, and would have undesirable resource implications. Concern was also expressed about lack of time for consultation with other agencies, and about how data zones would be grouped for release of data where data zones are too small for confidentiality reasons.

## **Highland**

The data zones originally proposed were revised by the local authority, because our methodology led to some villages being split (see East Ayrshire) and because the council wanted the zones to conform to the former district boundaries. These suggestions were accepted with minor modifications.

## **Inverclyde**

The data zones suggested were accepted, though a concern was expressed that data zones did not fit with ward boundaries.

## **Midlothian**

It was considered important that data zones fitted within ward boundaries, and a new set was produced by the council to improve this fit. We accepted these zones with minor alterations.

## **North Ayrshire**

Concern was expressed about problems of large rural zones incorporating unrelated settlements, but it was acknowledged that these arose from Output Area configurations and could not easily be rectified. Attention was also drawn to two islands, Little Cumbrae and Holy Island, that were originally unallocated to data zones, because they were not postcoded. These have now been allocated.

## **North Lanarkshire**

Concern was expressed about incompatibility with ward boundaries and with zone boundaries which split communities. Various changes to the zones were suggested and accepted except in a few cases where they did not conform with other criteria.

## **Perth & Kinross**

Several amendments were suggested, most of which were incorporated. These were usually intended to accord with planning areas or to split distinct settlements. In the remaining cases, matters were resolved so as to meet council needs while conforming with the other criteria.

## **Renfrewshire**

Several general criticisms of the proposed data zones were made, including a concern that many zones were not socially homogeneous and did not follow natural boundaries. Zones were felt to be too small, and to cut across urban / rural boundaries. They suggested redrawing the boundaries on the basis of local knowledge and priorities, but did not in fact do so, perhaps because of a shortage of time.

## **Scottish Borders**

A set of data zones was submitted which met most of our criteria and were felt to be most useful for local purposes. The main difference from our set was that rural and urban zones were separated, in some cases leading to zones with 'doughnut' shapes. Originally we had deliberately avoided such shapes but, after discussion with the Scottish Executive, we decided to allow them and hence accepted the new set of zones.

## **Shetland Islands**

Although we were in contact with this council, no substantive comments were received.

## **South Ayrshire**

This council raised concerns about the homogeneity of some of the data zones, although it was not able to make specific suggestions for changes. It favoured construction of the zones based on unit postcodes.

## **South Lanarkshire**

Changes were suggested on the basis of better fit to natural or community boundaries, involving reassignment of 199 out of 2444 Output Areas. These were accepted with minor changes.

## **Stirling**

The proposed data zones were accepted subject to a small number of changes, which were agreed with us.

## **West Dunbartonshire**

This council raised several issues concerning the exercise as a whole, such as how the data zones are to be updated, and how disclosure will be handled. It was also concerned that the data zones did not take account of proposed planning developments, something which it had been agreed not to do. After the end of the consultation exercise, a further letter was received from this council. This letter noted that West Dunbartonshire had responded according to the original fairly tight timetable for consultation. It went on to raise the issue that, although the consultation letter had implied that local authorities were not expected to undertake a comprehensive review, several authorities had done just that. Had this been known, West Dunbartonshire would have consulted more widely and identified and redrawn 'problem zones', resulting in a more useful set of boundaries. As it stands, the Council claims that the data zones are no longer comparable between authorities, significantly reducing the value of the exercise.

## **General comments**

The consultation exercise seems to have been immensely valuable. Almost all authorities were able to suggest changes, minor or major, which from their points of view improved the system significantly. It also caused us to re-evaluate some of the decisions we had made in abstract in the light of cases made for change. This process resulted in accepting zones with a household population of slightly less than 500 or greater than 1000 where there was a strong case for doing so, in accepting 'doughnut-shape' zones, which had originally been inadmissible, and in taking more account of settlement boundaries. We also accepted changes designed to fit better with other

areal units considered to be important in particular areas: these included former county and district boundaries, ward boundaries, and neighbourhood units already used for planning purposes in some authorities. Although we have some sympathy with West Dunbartonshire (and perhaps other authorities), who might have suggested more far-reaching changes if they had thought they would be accepted, we do not feel this is a reason for rejecting changes which will make the zones more useful for other authorities.

## **Conclusion**

The task of defining data zones across Scotland proceeded reasonably smoothly and satisfactorily. The basic structure agreed in the previous project proved to be workable, although several changes in ground rules were agreed following cases being made by various local authorities. Criticisms made of the original data zones issued for consultation were related to decisions made in consultation with the Scottish Executive. The most important criticisms related to the effects of using primary school catchment areas as the starting point for zone definition, as opposed to the lack of emphasis given to settlement boundaries, and the avoidance of ‘doughnut’ shapes. Data zones were also criticised by some authorities as being too large or too small; this was probably inevitable given the decision that sizes should be comparable across Scotland.

Apart from the ‘ground rules’ that had to be modified, our method worked reasonably well. Within our constraints, there are many ways in which data zones can be constructed in any area, and without local knowledge, we could only select what appeared to be the best. We are not surprised or dismayed that most councils were able to suggest improvements. Even where councils suggested substantial revisions, we believe that our initial proposals were a valuable starting point. We thought it was important to accept local councils’ suggestions where possible, on the grounds that the data zones should be as useful as possible to the community as a whole, and local authorities are likely to be among the most frequent users of Neighbourhood Statistics.

In conclusion, we would like to thank staff in the local authorities for their work in reviewing and revising our proposals, usually in a friendly and constructive manner, and Robert Williams of the Scottish Executive for his guidance in the project.

## Figure 1 Instructions for constructing data zones based on 2001 data

### Prepare data files:

ArcInfo is used to prepare the data files (any ESRI ArcGIS packages can be used). Suppose the shape file is already loaded into ArcInfo.

1. Generate OA identifiers (oa\_id) for 2001 census output areas (OAs).  
Choose the OA boundary shape file for a council area. Open the attribute table (*right click the shape file layer and click Open Attribute Table*). Add field 'oa\_id' (*left click Options in the bottom bar and click Add Field, defining 'oa\_id' as integer*) and calculate values by loading:the VB macro:  
GetOA\_ID.cal
2. Create geometrical centroids for OAs.
  - 2.1 Open the attribute table. Add field 'x' and field 'y' (*define 'x' and 'y' as integer*). Calculate values by loading separately:  
Cal\_x.cal  
Cal\_y.cal
  - 2.2 Export the attribute data to a dbf file (*right click the shape file layer and click Open Attribute Table, click options in the bottom bar, click Export, provide a file name, click OK, when prompted 'Do you want to add the table to ArcMap?', click yes*). From the drop-down menu Tools, click on 'add XY data'. Select the centroids dbf file and assign x to field 'x' and y to field 'y'. Click OK to produce a point shape file for centroids of OAs.
3. Add population and Townsend index data.
  - 3.1 Join with the data file (sct\_pop\_town.csv) containing population and Townsend index for OAs. There are two types of population: population in households and population in communal establishments. We use population in households to define population size. Note that population in households is named as 'hhpop'.
4. Create pseudo-school catchment area boundary and assign OAs to school\_ids.
  - 4.1 Join the centroid shape file with the school catchment shape file based on location (this is a point-in-polygon operation). The result is a point shape file representing centroid points and having school attributes attached to each OA.
  - 4.2 Join back the oa\_to\_school shape file to the OA boundary shape file by 'Tag'.
  - 4.3 Dissolve by school ID to get a pseudo-school catchment boundary shape file.
  - 4.4 Then convert it into a coverage using ArcTools.
  - 4.5 Add the pseudo-school catchment boundary coverage onto ArcMap. Open its attribute table, then summarise the attribute table by school\_ID.
  - 4.6 Open the summary table, sort by count of school\_ID. If the count is more than 1, it means that the pseudo-school catchment is multi-extent: we need to examine the OA boundaries visually and amend or remove OAs to make the pseudo-school catchment a single polygon.
  - 4.7 Export the attribute table.
5. Open the attribute table into Excel or another spreadsheet package to delete unwanted fields. Save the data file in the format:

**File 1 (\*.dat): oa\_id, x, y, population, townsend, school\_id**

6. Generate a contiguity matrix. Convert the OA shape file into a coverage file. Start the Arc workstation. Run the seg2con.aml (&r seg2con) to produce the contiguity matrix file. Oa\_id should be used to provide polygon IDs.

**File 2 (\*.con): contiguity file**

**Run the program:**

Start the program by typing:

```
Datazone datafile contiguity_file out_popfile out_zoneid_file  
min_population max_population weight_for_compactness  
weight_for_homogeneity
```

(e.g. for Fife council, type: datazone fife.dat fife.con fife.pop zoneid.txt 500  
1000 2 1)

The results file has the format:

**oa\_id, zone\_id**

Join this file to the OA shape file and dissolve by zone\_id into data zone shape file.

**Quality assurance:**

1. Check if there are zones made up of non-contiguous polygons.
  - i. Convert the datazone shape file into a coverage
  - ii. Summarise by zone\_id
  - iii. Check if zones with count larger than 1 are multi-extent.
2. Population size check. For zones over 1000 or under 500 people check if the population can be brought between 500 and 1000 by swapping one or two OAs with neighbouring zones.
3. Check shape compactness:
  - use the coverage just created to add a new field 'ratio' and calculate  $\text{ratio} = \text{squared perimeter} / \text{area}$
  - check the zones with large scores.
4. Check relationship with other boundaries, such as settlement boundaries.
5. Check relationship with environmental features (e.g. roads, rivers, railways) by comparing with Digimap 1:50 k colour raster (topographical maps). Swap OAs between data zones where necessary.

**Final results:**

- 1 Data zone shape file, containing: data zone ids, population in households, population in communal establishments
- 2 Look-up table linking OAs to data zones
- 3 Data zone summary, containing local authority code, population in households, population in communal establishments, average Townsend score

## **Figure 2 Amalgamation program used to create data zones**

This program is written in FORTRAN. The program procedure is as follows:

Step 1 Group OAs by primary school catchment and calculate sum of household populations for each catchment. The set of OAs with centroids in the primary school catchment area is called a pseudo-catchment.

Step 2 Create data zones school by school

Step 3 If the population of a pseudo-catchment is below 1000, then the OAs in the pseudo-catchment are kept as one zone. Go back to step 2. Otherwise, go to step 4.

Step 4 If the population of a pseudo-catchment is over 1000, set the number of zones = pseudo-catchment population / 1000, rounded up; the population size = pseudo-catchment population / number of zones.

Step 5 Select the most southwesterly OA to start with.

Step 6 Check if sum of OA populations > the population size. If so, stop and keep OAs as one zone and go to step 5. If not, go to Step 7.

Step 7 Identify OAs contiguous to the OAs in the zone.

Step 8 Calculate Euclidean distances from the selected OA to the contiguous OAs.

Step 9 Calculate 'distances' between Townsend scores of contiguous OAs and the selected OA.

Step 10 Rank geographical distances and Townsend 'distances' separately

Step 11 Construct a combined score by adding ranks of physical distances (double weighted) and Townsend 'distances'.

Step 12 Pick the OA with the lowest score. If the sum of OA populations is below the population size, keep the OA in the zone and go to step 7; if not, go to step 13.

Step 13 Remove the OA with the lowest score from the list of contiguous OAs, and go back to step 12.

Step 14 Examine zone by zone whether the population threshold is met.

Step 15 If zone population is below 500, join to another contiguous zone with population below 500, or join to another contiguous OA from a second zone, provided first that the removal of the OA does not split the second zone, and second that the population of the second zone remains at 500 or above.

Step 16 If population of a zone is over 1000, remove an OA from the zone, provided first that the removal of the OA does not split the zone, and second that the population remains at 500 or above.

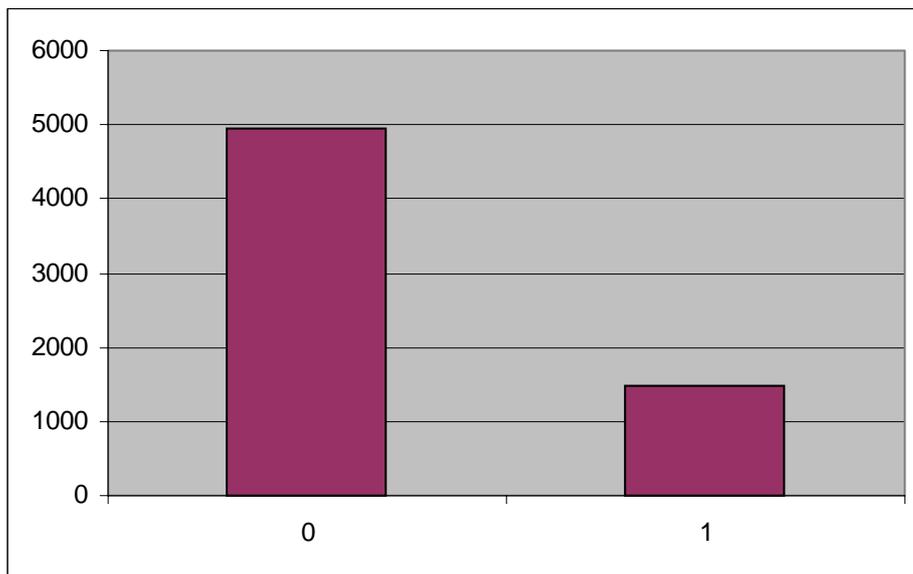
**Figure 3 Number and size of data zones by local authority**

Local authority	Number of zones	Average household population	Average total population
Aberdeen City	267	773.47	794.48
Aberdeenshire	301	744.11	753.72
Angus	142	749.61	763.38
Argyll & Bute	122	716.42	748.41
Clackmannanshire	64	736.31	751.20
Dumfries & Galloway	193	755.35	765.62
Dundee City	179	796.09	813.76
East Ayrshire	154	770.13	780.75
East Dunbartonshire	127	842.24	852.31
East Lothian	120	741.30	750.73
East Renfrewshire	120	738.73	744.26
Edinburgh, City of	549	797.44	817.17
Eilean Siar	36	726.81	736.17
Falkirk	197	728.21	737.01
Fife	453	757.15	771.37
Glasgow City	694	815.52	832.66
Highland	292	703.26	715.46
Inverclyde	110	755.05	765.48
Midlothian	112	714.64	722.69
Moray	116	731.65	749.48
North Ayrshire	179	750.99	758.75
North Lanarkshire	418	761.06	768.10
Orkney Islands	27	704.96	712.78
Perth & Kinross	175	747.75	771.14
Renfrewshire	214	797.70	807.79
Scottish Borders	130	810.63	821.26
Shetland Islands	30	722.60	732.93
South Ayrshire	147	751.28	762.56
South Lanarkshire	398	750.78	759.34
Stirling	110	761.18	783.75
West Dunbartonshire	118	784.96	791.34
West Lothian	211	746.83	752.20

**Figure 4 Relationship of data zones to primary school catchment areas**

	Frequency	Percentage
Same Catchment (0)	4945	77.1
Different Catchment (1)	1467	22.9
Total	6412	100

The table and bar chart exclude Shetland, Orkney and Eilean Siar where we do not have digital school catchment boundaries and Voronoi polygons were not used to approximate school catchment boundaries.

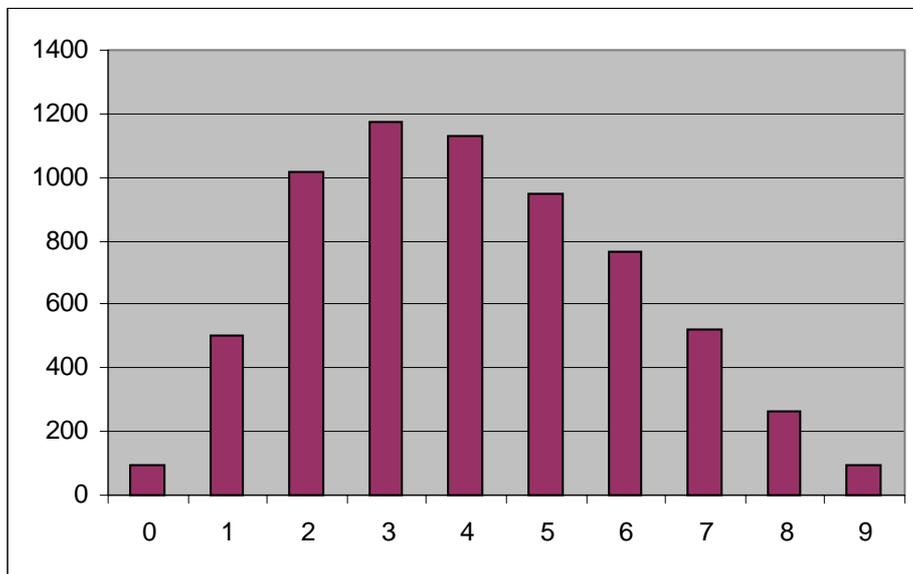


The left-hand column shows the frequency of data zones that are wholly within one primary school catchment area (as modified to fit Output Area boundaries). The right-hand column shows the frequency of data zones that overlap these boundaries.

**Figure 5a Social homogeneity of data zones as measured by the modified Townsend index**

	Frequency	%
0	97	1.49
1	503	7.73
2	1015	15.60
3	1171	18.00
4	1130	17.37
5	949	14.59
6	767	11.79
7	518	7.96
8	262	4.03
9	93	1.43
Total	6505	100

Townsend scores (modified to exclude student unemployment) were computed for each Output Area and divided into deciles. The left column represents the maximum difference between Output Areas within the same data zone. If all Output Areas within a data zone have modified Townsend scores in the same decile, the value is 0; if the data zone includes Output Areas in the first decile and in the tenth decile, the value is 9.



The bars represent the frequency of data zones of different degrees of homogeneity. The codes run from 0 for the most homogeneous data zones to 9 for the least homogeneous.

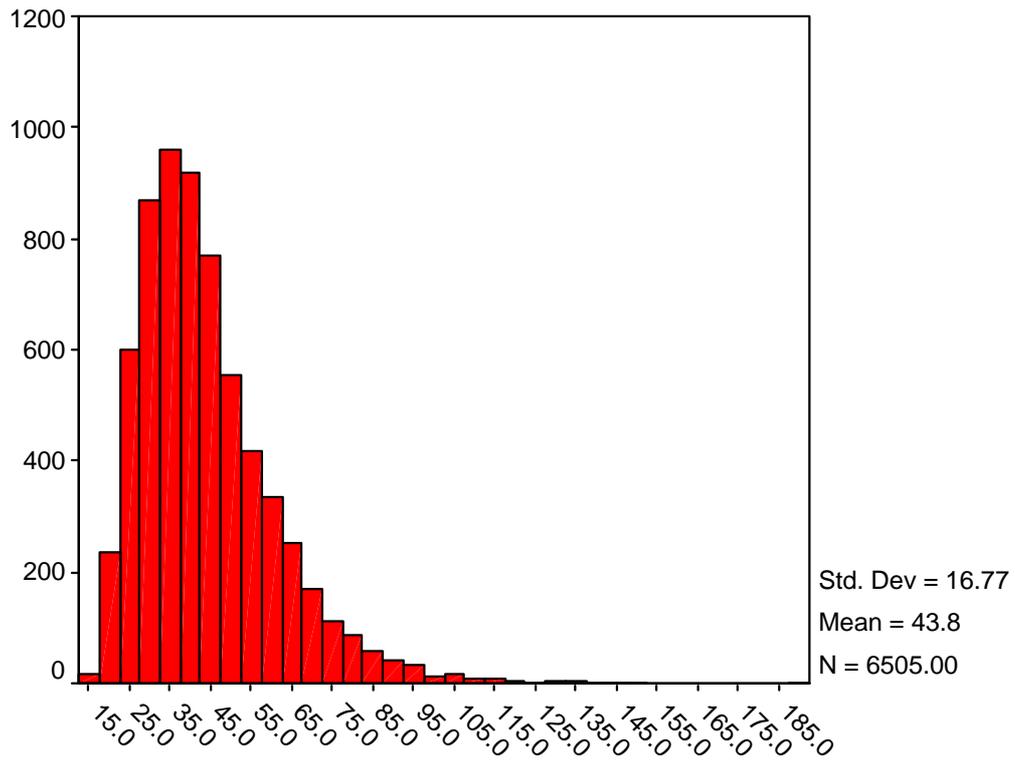
**Figure 5b Electoral wards: decile distribution**

Difference	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1	1	0.08	1	0.08
2	10	0.82	11	0.90
3	24	1.96	35	2.86
4	44	3.60	79	6.46
5	91	7.45	170	13.91
6	138	11.29	308	25.20
7	229	18.74	537	43.94
8	316	25.86	853	69.80
9	369	30.20	1222	100.00

**Figure 6 Compactness of data zones for local authorities**

<b>Name</b>	<b>council code</b>	<b>number of zones</b>	<b>average</b>	<b>standard deviation</b>
Aberdeen City	100	267	42.68	15.15
Aberdeenshire	110	301	49.75	19.14
Angus	120	142	45.12	14.04
Argyll & Bute	130	122	52.12	21.33
Clackmannanshire	150	64	45.17	14.07
Dumfries & Galloway	170	193	48.52	17.15
Dundee City	180	179	38.32	13.55
East Ayrshire	190	154	47.88	21.23
East Dunbartonshire	200	127	42.22	12.73
East Lothian	210	120	45.29	16.06
East Renfrewshire	220	120	42.36	13.30
Edinburgh, City of	230	549	40.43	13.27
Eilean Siar	235	36	58.14	27.57
Falkirk	240	197	39.78	15.76
Fife	250	453	44.70	16.36
Glasgow City	260	694	41.05	14.09
Highland	270	292	46.29	21.11
Inverclyde	280	110	45.46	17.23
Midlothian	290	112	37.08	13.57
Moray	300	116	46.19	15.38
North Ayrshire	310	179	39.62	16.31
North Lanarkshire	320	418	40.68	14.36
Orkney Islands	330	27	53.15	22.06
Perth & Kinross	340	175	51.55	17.48
Renfrewshire	350	214	37.69	14.66
Scottish Borders	355	130	47.35	16.39
Shetland Islands	360	30	77.34	30.41
South Ayrshire	370	147	45.11	17.17
South Lanarkshire	380	398	46.37	16.58
Stirling	390	110	49.24	15.22
West Dunbartonshire	395	118	44.01	15.96
West Lothian	400	211	37.22	13.42

The average and standard deviation referred to above are calculated from the ratio of perimeter (in metres) squared divided by area (in square metres), used as a measure of compactness. This ratio is computed for each data zone in a local authority, and averages and standard deviations are calculated in the usual way.



COMPACT

## Appendix 1 Letter to Chief Executives

### SCOTTISH EXECUTIVE Finance and Central Services Department

Chief Executive Office

Central Statistics Unit  
St Andrew's House  
Regent Road  
Edinburgh EH1 3DG

Telephone: 0131-244 0443  
Fax: 0131-244 0335  
robert.williams@scotland.gsi.gov.uk  
<http://www.scotland.gov.uk>

Your ref:  
Our ref: H/4/8

**XXX September 2003**

---

Dear Sir or Madam

#### **SCOTTISH NEIGHBOURHOOD STATISTICS: DATA ZONES BASED ON 2001 CENSUS**

The Scottish Executive is setting up a Scottish Neighbourhood Statistics programme intended to make available small-area data of many different kinds across Scotland. This work has involved creating a set of pre-defined data zones for the whole of Scotland in order to make a range of neighbourhood statistics available at the local level, subject to appropriate confidentiality constraints. The University of St Andrews is creating these data zones on behalf of the Scottish Executive.

You may remember participating in a similar exercise last year at which time the methodology for creating data zones was agreed. A prototype set of data zones based on 1991 Census output areas was also created. At that time it became clear that in some local authorities changes since 1991 made the original set of zones inappropriate and the Scottish Executive agreed to update the data zones using information from the 2001 Census. Now that small-area data are available for the 2001 Census, we have recreated data zones to take account of these changes and other points made by local authorities.

The data zones are aggregates of unit postcodes and census output areas and nest within Local Authority boundaries. They are designed to have a household population size between about 500 and 1000, to have a reasonably compact shape, to observe natural boundaries, and to be relatively socially homogeneous. It is not possible to meet all these constraints, and the zones produced are intended to be a good compromise between the various relevant factors.

Because of the number of data zones, we have sent you the zone boundary data for your authority on the enclosed CD rather than in paper form. It is designed to be read using ESRI software, such as ArcView or ArcInfo. Last year the University of St Andrews sent you another CD containing ESRI's Map Explorer software, which can be used to read the data if your GIS is not compatible with ESRI products. This can be used again. If you are not able to access the data via GIS, the University of St Andrews can supply .pdf, .gif or .jpg images of the data.

The enclosed CD contains files representing the boundaries of our suggested data zones, Excel files listing the 2001 census population for each data zone and a lookup table relating data zones to 2001 Census output areas. It should be possible, if you wish, to plot the data zones onto a digital backdrop of topographical data that you will have been supplied with by the Ordnance Survey.

The methodology has been agreed, but it is possible to make adjustments to the data zone boundaries based on your local knowledge and we hope an appropriate person within your organisation will be able to look at the proposed data zones and their boundaries. **Comments and suggested adjustments are sought by 20 October 2003, and should be sent to Robin Flowerdew at the University of St Andrews.**

Scottish Neighbourhood Statistics can be found at [www.sns.gov.uk](http://www.sns.gov.uk) and information on the datazone project can be found at <http://www.scotland.gov.uk/stats/neighbours/tables/geography.asp>.

Thank you in advance for your help.

Yours faithfully

Robert Williams  
Scottish Executive

And

Robin Flowerdew  
University of St Andrews

**Comments and suggested adjustments should be sent to:**

**Professor Robin Flowerdew  
University of St Andrews  
School of Geography & Geosciences  
Irvine Building  
North Street  
St Andrews  
Fife KY16 9AL  
Tel: 01334 463853  
Fax: 01334 463949  
Email: rf15@st-and.ac.uk**

## Appendix 2 Follow-up letter

Office of the Permanent Secretary

Local Authority Contacts  
Data Zone Development  
Scottish Neighbourhood Statistics Project

Central Statistics Unit  
St Andrew's House  
Regent Road  
Edinburgh EH1 3DG

Telephone: 0131-244 7310  
Fax: 0131-244 0335  
Peter.whitehouse@scotland.gsi.gov.uk  
<http://www.scotland.gov.uk>

21 Oct 03

### Purpose

1. To note the revised timetable for responding to the consultation exercise on developing Data Zones. The Scottish Executive has further extended the original timetable and you will wish to note that revisions to the draft Data Zones received by St Andrew's **on or before 14<sup>th</sup> November 03** will be considered.

### Timing

2. Urgent. This is for immediate attention.

### Background

3. The development of Data Zones has been progressed over the last 18 months and can be seen in three stages.

- a) discussion with SNS Development Group leading to presentation of initial ideas from LAs.
- b) development of these ideas into a single approach and the production of 'draft' Data Zones based in 1991 Census Output Areas.
- c) production of 2001 based Data Zones using the common methodology delivered by stages a) and b).

4. The work is currently in the final stages of c) and LAs have been invited to consider whether there are any minor adjustments that need to be made to the 2001 based Data Zones which will allow them to represent better the local area.

5. The purpose of the final revisions by LAs is not to have a root and branch change of the Data Zones, but to allow specific local knowledge to 'fine tune' the Data Zones. The initial timescale of a response by 20<sup>th</sup> October reflected the expectation that this task was relatively minor.

## Discussion

6. The Scottish Executive is now aware that some LAs are unhappy with the timescale and the constraints within which the Data Zones need to be agreed. The SE has therefore concluded the following

- a) the basic methodology for constructing the Data Zones has been agreed and it is not appropriate at this late stage to reconstruct Data Zones using alternative threshold and boundary data.
- b) the population thresholds of 500 and 1000 (household population) should be regarded as firm for the vast majority of Data Zones. However if there are clear reasons why for a specific Data Zone a small reduction to an absolute minimum of 475 people (household population) or an increase to an absolute maximum of 1,100 people (household population) would give a significantly improved area, then this can be considered by the LA.
- c) the Data Zones methodology requires them to nest within LA boundaries. The requirement to nest within Ward boundaries and other boundaries was considered during the earlier stages of the Data Zone work and was dismissed as either being subject to frequent revision, not nesting into LA boundaries, not consistent with Census Output Area (COAs) boundaries or not present across Scotland. However, if in a specific area of a LA there is a natural or constructed boundary which if respected would make the Data Zone a better descriptor of the local area, then LAs are able to swap around some relevant COAs to accommodate this local knowledge.

7. The deadline for returning amended Data Zones to St Andrew's will be **14<sup>th</sup> November**. This is the final deadline and in order that the SNS work can progress, only amendments received by St Andrew's **on or before 14<sup>th</sup> November will be considered**. It is expected that the vast majority of Data Zones will be accepted as presented in the September 03 versions.

8. In order to allow revisions to be considered, each Output Area that the LA wishes to re-allocate should be listed alongside its current Data Zone and also the preferred Data Zone into which the LA would want it to be placed.

9. Some LAs have called for a larger geography to be made available via the SNS system. In order to accommodate this, the SE will implement a consultation exercise during 2004 to deliver an 'Intermediate Geography'. Further details on this will be forthcoming in early 2004.

## Conclusion

10. LAs can return proposals for adjusting the draft Data Zones to reach St Andrew's **on or before 14<sup>th</sup> November**. Returns received

after this date will not be considered. The basic methodology for building Data Zones will be respected, but specific adjustments to the groupings of COAs will be considered where local knowledge suggests a more appropriate construct. It is expected that the vast majority of September 03 Data Zones will be accepted by LAs and that LAs will only be identifying a small minority of Data Zones for re-working.

11. I hope you will agree that significant progress has been made over the last 18 months and that the Data Zones and wider SNS geography will have significant benefits to those developing policy and providing services at a local area level. Thank you for your contribution in progressing this work and for your continued involvement in this important matter.

Pete Whitehouse  
Senior Statistician  
Scottish Executive  
21 October 03.

## References

Alvanides S, Openshaw S and Rees P (2002) Designing your own geographies, in Rees P, Martin D and Williamson P (eds) *The census data system* (Wiley) 47-65

Martin D (2002) Geography for the 2001 Census in England and Wales *Population Trends* 108, 7-15

Townsend P, Phillimore P and Beatty A (1988) *Health and deprivation: inequality and the North* (Croom Helm)

**Schedule 6 Licence for the  
use of Ordnance Survey  
Derived Data**

The following licence to be used by Licensees wishing to pass on derived data products to third parties as described in Schedule 1 Section 6.

This Data licence is designed to be an end user **break seal** agreement. It may also be used as a **click to accept terms and conditions** for an online data download.

For **break seal** agreements – your opening of this package will be deemed to be your acceptance of the following terms. If you do not wish to accept the terms, return this package within fourteen (14) days unopened to us or our Agent or other party from whom you obtained this package.

For **click to accept terms and conditions** agreements – before you click on the **accept** button at the end of this document and download the data, you must carefully read the terms and conditions of this licence. By clicking on the **accept** button and by accessing the data you will be deemed to be bound by the terms of this licence. If you do not wish to be bound then please click the **cancel** button now.

### **Licence**

We (Scottish Executive) grant you a non-exclusive non-transferable licence (without the right to sublicense) to copy and use the Data which is derived from Ordnance Survey data in accordance with the terms of this licence agreement.

You may only use the Data for your own internal business use, that is use of Data for the internal administration and operation of your business and not for any commercial purpose, and not for financial profit or gain. Financial gain would include any profit whether direct or indirect, or benefit from the use or publication of the Data in any form.

### **Intellectual Property Rights**

Intellectual Property Rights in the Data are owned by the Crown and us. You shall not have any rights or interests in the Data other than as described in this licence.

You must ensure that you protect and do not interfere with any trade names and Trade Marks that are in or accompany the Data.

All copies of the Data in whatever form must contain the following acknowledgement “© Crown Copyright. All rights reserved [year]”.

### **Restrictions**

You may only use the Data for your internal business use. You must not sublicense, lend, transfer or otherwise dispose of the Data or this licence.

You must not sublicense, lend, sell, lease, assign, transfer or otherwise dispose of the Data or this licence.

You must not modify, alter, decompile, reverse engineer, or disassemble the Data beyond that which is necessary to allow the use of the Data within your system.

### **Promises**

We promise that we have the power to grant this licence. However, you acknowledge that the data has not been prepared to meet your individual requirements and therefore it is provided to you on an **as is** basis. It is your responsibility to ensure that the data is fit for your intended use. To the fullest extent permitted by law we exclude any conditions or terms that may be implied by law.

We shall not in any event be liable for any indirect, special, consequential, or incidental losses or for loss of data, loss of profits, loss of or interruption to business whether arising in tort (including negligence) contract or otherwise. We do not attempt to exclude any liability that cannot be excluded. Excepting the above, our total and aggregate liability to you in contract, tort (including negligence) or otherwise will not at any time exceed an amount equal to the licence fee (if applicable) paid by you.

**Term**

You are licensed to use the data until 3<sup>rd</sup> February 2013. This licence will immediately terminate on termination of the licence between Ordnance Survey and Scottish Executive, if such termination occurs prior to the end of the Term.

**Termination**

We may terminate this licence at any time if any of the following events arises:

You breach any of the terms of this licence;

You become insolvent or bankrupt or cease to do business.

In the event of termination of this licence, you shall cease to use the Data and if we request, you shall immediately either return or destroy the Data. You will provide us with written confirmation of the action you have taken.

**Rights of Third Parties**

Furthermore, we both agree that Ordnance Survey and the Controller of Her Majesty's Stationery Office shall benefit from the Contracts (Rights of Third Parties) Act 1999 so that Ordnance Survey can enforce its rights directly against you.

**Law**

This will be governed by and construed in accordance with English law and both parties accept the exclusive jurisdiction of the English courts. If any part of this licence is held to be invalid, unenforceable, or illegal, we both agree that the remainder of the licence will stand.