

# **Revising Small Area Statistics Geographies – Data Zones and Intermediate Zones: Consultation Document**

July 2024

## Contents

<b>Key Points</b> .....	<b>2</b>
<b>Introduction</b> .....	<b>3</b>
<b>What are Data Zones and Intermediate Zones?</b> .....	<b>4</b>
Data Zones .....	4
Intermediate Zones .....	5
<b>Why we need to revise Data Zones and Intermediate Zones</b> .....	<b>5</b>
Population change over time .....	5
Changes to higher level boundaries .....	6
Changes to lower-level boundaries .....	6
<b>Methodology</b> .....	<b>7</b>
Boundaries .....	7
Centroids .....	8
Population and household totals .....	8
Names and Codes .....	9
<b>Summary of changes</b> .....	<b>9</b>
<b>Consultation questions</b> .....	<b>9</b>
<b>Responding to this consultation</b> .....	<b>10</b>
Handling your response .....	11
Next steps in the process.....	11
Comments and complaints .....	11
Scottish Government consultation process.....	12

# Key Points

- Data Zones and Intermediate Zones are small area geographies used in the production of official statistics in Scotland.
- Data Zones serve two main functions in the production of official statistics. Firstly, they provide a standard, stable geography for the production and publication of statistics at a local area on a wide range of topics. Secondly, aggregations of Data Zones can be used to provide estimates for higher level geographies where official statistics might not otherwise be available.
- Intermediate Zones are similar to Data Zones but cover a larger area and population. They are intended to be used to produce local area statistics where statistics cannot be published at Data Zone level due to risk of disclosing sensitive information.
- We need to revise Data Zones and Intermediate Zones because of changes to population and other geographical boundaries since they were last revised in 2014. In particular, the proposed changes have been produced to align with new population data from the 2022 Census.
- The purpose of this consultation is to seek feedback from users on our proposed revisions to Data Zones and Intermediate Zones.
- Responses to this consultation are invited by 1<sup>st</sup> October 2024. Please respond to this consultation using the [Scottish Government's consultation hub](#).
- We aim to publish finalised Data Zones and Intermediate Zones as soon as possible following the conclusion of this consultation, and before the end of 2024.

# Introduction

Data Zones and Intermediate Zones are small area geographies used in the production of official statistics in Scotland. They were first introduced in 2004 and revised in 2014. The purpose of this consultation is to seek feedback from users on proposals to update these geographies to ensure that they remain fit for purpose.

Data Zones are designed so that they are roughly comparable in population size, with around 500 to 1,000 people. There are currently 6,796 Data Zones in Scotland. They serve two main functions in the production of official statistics. Firstly, they provide a standard, stable geography for the production and publication of statistics at a local area on a wide range of topics. Secondly, aggregations of Data Zones can be used to provide estimates for higher level geographies where official statistics might not otherwise be available.

Intermediate Zones are similar to Data Zones but cover a larger area and population. They are designed to have populations of between 2,500 and 6,000, and there are currently 1,279 Intermediate Zones. They are intended to be used to produce local area statistics where statistics cannot be published at Data Zone level due to the risk of disclosing sensitive information.

We need to revise Data Zones and Intermediate Zones because of changes to population and other geographical boundaries since they were last revised in 2014 following the 2011 Census. In particular, the proposed changes have been produced to align with new population data from the 2022 Census.

The purpose of this consultation is to seek feedback from users on our proposed revisions to Data Zones and Intermediate Zones.

The proposed Data Zones and Intermediate Zones are available to view at [Proposed statistical geographies \(arcgis.com\)](#) and download at [Data Zones and Intermediate Zones consultation - Overview \(arcgis.com\)](#).

Responses to this consultation are invited by 1<sup>st</sup> October 2024. Please respond to this consultation using the [Scottish Government's consultation hub](#). More information on how to contribute to this consultation is included at the end of this document.

We aim to publish finalised Data Zones and Intermediate Zones as soon as possible following the conclusion of this consultation, and before the end of 2024.

# What are Data Zones and Intermediate Zones?

Data Zones and Intermediate Zones are small area geographies used in the production of official statistics. This section describes what Data Zones and Intermediate Zones are, and how they are used.

## Data Zones

Data Zones are the key geography used for producing small area statistics in Scotland. They were first made available for use in 2004 following the publication of results from the 2001 Census and were produced to promote the production of local area statistics. They were revised in 2014, following the release of 2011 Census results.

Data Zones serve two key functions in the production of statistics. Firstly, they provide a standard, stable geography that can be used to publish statistics at a local level across a wide range of topics. They are designed to be large enough that statistics can be presented accurately without fear of disclosure or personal information. They are also designed to be small enough that they can be used to represent communities.

Secondly, Data Zones are used as a building block to construct statistics for higher geographies. Aggregations of Data Zones are often used to approximate larger areas of geography for statistics which might not otherwise be available. To support this, National Records of Scotland include Data Zones in their [Scottish Postcode Directory](#), which provides a 'lookup' to higher geographies.

The use of Data Zones as statistical building blocks supports the implementation of the [Government Statistical Service Geography Policy](#). For example, this policy states that statistics for all geographies should be built by first recording information at 'building block' level and then 'best-fitting' these building blocks to higher geographies, rather than recording or assigning information at a higher level directly. In Scotland, Data Zones, as well as Census Output Areas and Intermediate Zones, are the designated statistical building blocks under this policy.

Data Zones are built by combining Census Output Areas and were initially designed to satisfy a number of constraints. These are:

- Approximate equality of population, so that each Data Zone has a population of 500 to 1000 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.

In addition, Data Zones are designed so that they nest within Local authority boundaries. In practice it is impossible to satisfy all these criteria, so Data Zones must find a compromise between these criteria.

## **Intermediate Zones**

Intermediate Zones are similar to Data Zones but cover a larger population and area. They are created by combining Data Zones to produce areas with a population of between 2,500 and 6,000. As for Data Zones, Intermediate Zones are designed so that they nest exactly into local authority boundaries.

Intermediate Zones are intended to be used to produce local area statistics where statistics cannot be published at Data Zone level due to the risk of disclosing sensitive information. The larger population of Intermediate Zones means that statistics are less likely to require disclosure control.

# **Why we need to revise Data Zones and Intermediate Zones**

Data Zones and Intermediate Zones are designed to be stable geographies. There is a need, however, to revise them periodically to ensure they remain fit for purpose.

There are a number of reasons why Data Zones and Intermediate Zones are periodically updated. These reasons are discussed in this section.

## **Population change over time**

The main reason why we need to revise Data Zones and Intermediate Zones is to account for population change. These geographies are designed to have roughly similar populations. The current versions used data from the 2011 census to ensure this requirement is satisfied.

Over time, however, changes in the population can cause some Data Zones to fall below or exceed the population thresholds. This can be due to new houses being built, old houses being demolished, or other changes in population distribution across Scotland. To account for these population changes, Data Zones and Intermediate Zones must periodically be reviewed using more current data.

In particular, revisions are timed to coincide with the [Scottish Census](#). Data Zones and Intermediate Zones were first produced in 2004 based on 2001 Census results. They were then revised in 2014 so that they were based on 2011 Census data. The purpose of the proposals outlined in this consultation is to update Data Zones and Intermediate Zones so that they are based on recently published Census 2022 data.

## **Changes to higher level boundaries**

Data Zones and Intermediate Zones are also revised to take into account changes to other geographic boundaries, most notably local authority boundaries. This is because Data Zones are designed so that aggregations of Data Zones produce exact matches of local authorities.

Local authority boundaries rarely change, however, when they do it means that aggregations of Data Zones no longer match Lboundaries exactly, so statistics produced from Data Zones become approximations based on best-fit aggregations.

This can also introduce disclosure risk. This is because 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.

There have been two boundary changes to local authorities since 2011. A small change, which did not affect any population, was made between [Fife and Perth and Kinross at Keltlybridge and Fife Environmental Energy Park, Westfield](#) in February 2018. A slightly larger change, which did affect some population, occurred between [Glasgow City and North Lanarkshire at Cardowan by Steps](#) in April 2019.

## **Changes to lower-level boundaries**

Data Zones are produced by aggregating Census Output Areas. Output Areas in turn are produced by aggregating 'frozen' Census postcodes. Changes in postcode boundaries over time can erode this link between postcodes and Data Zones, which means postcodes will be aggregated up to Data Zone level on a best-fit basis.

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 National Records of Scotland (NRS) produce postcode boundaries based on the location of address points with the same postcode. A 'frozen' set of Census postcode boundaries is used to create Data Zones. NRS publish updated postcode boundaries every six months, but the Data Zone boundaries remain fixed.

There are two issues associated with postcode drift that affect Data Zones. Firstly, as mentioned above, 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, which remain fixed.

The second 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.

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. This can introduce inaccuracy into statistics based on postcodes. Further information on the NRS postcode products is available from the [NRS website](#).

## Methodology

This section outlines the methodology we have used to update Data Zones and Intermediate Zones.

### Boundaries

The method which was used to update 2001 Data Zone boundaries following the 2011 Census has been replicated for this Census 2022 update. The aim for this update was to apply Census 2022 data, re-align to local authority boundaries (as at Census Day 2022), and account for changes in population, while keeping boundaries relatively comparable to those created for 2011. The criteria used in the definition of 2022 Data Zones were:

- That they be built up from 2022 Census Output Areas.
- Maintain approximately equal household resident populations of 500 to 1,000 people, with an absolute minimum of 375 people and a maximum of 1,125.
- Be a continuous area without multiple extents, unless caused by water (e.g. Data Zones including islands).
- Where possible, changes from 2011 occur within Intermediate Zone boundaries.
- Maintain a compactness of shape.

Initially, a first draft of boundaries was created by aggregating 2022 Census Output Areas to 2011 Data Zones, based on the location of the Output Area Population Weighted Centroids. The population (defined as household residents) was summed for each 'best-fit' Data Zone to assess whether it fell within the required population thresholds. Data Zones which had seen a decrease in population (e.g. as the result of housing demolition) and fell below the minimum population of 375 were merged with the neighbouring Data Zone with the longest shared boundary within the same



Intermediate Zone. Conversely, Data Zones which had seen an increase in population to values greater than 1,125 (e.g. as the result of a new housing development) were split into one or more new Data Zones.

After the split and merge process was complete, additional changes were applied to reconfigure any multi-extent Data Zones which were not caused by water. This occurred where Census Output Areas had changed significantly from 2011 and resulted in some best-fit 2022 Data Zones being composed of two parts that do not touch. In most instances, the smaller non-contiguous part of the Data Zone (usually consisting of a single Output Area) was absorbed into a neighbouring Data Zone. In addition, we had already received feedback from some local authorities on changes to make to Data Zone boundaries, which was considered and applied where possible when designing the proposed 2022 Data Zones.

A similar process was followed for Intermediate Zones. The first step was to best-fit the proposed 2022 Data Zones onto the 2011 Intermediate Zone boundaries using the Data Zone Population Weighted Centroids. Then, any 'best-fit' 2022 Intermediate Zones which fell below the population threshold of 2,125 household residents, or above 6,375 household residents, were reconfigured. In addition, any best-fit 2022 Intermediate Zones which had become multi-extent (excluding those caused by islands) were reconfigured.

## **Centroids**

Once the proposed boundaries had been created, centroids were calculated. This is not the geometric middle of the Data Zone or Intermediate Zone, but a point that represents the population weighted centre. The main use of a centroid is to allocate a Data Zone or Intermediate Zone to a higher-level geography. The 'median method' which was used in 2011 has also been used to calculate the proposed 2022 Data Zones and Intermediate Zones centroids. The method is as follows:

1. The median easting and northing coordinate pair for all Census Output Areas within the Data Zone is calculated, weighted by the household residents population, giving a notional centroid of the Data Zone.
2. The distance from each of the Census Output Area centroids to the median easting/northing is calculated. The Census Output Area centroid with the shortest distance to the median is chosen as the centroid of the Data Zone.

The same process is followed for Intermediate Zones, using the Data Zone centroids as the input (i.e. the Data Zone centroid closest to the median easting/northing is chosen as the Intermediate Zone centroid).

## **Population and household totals**

The total number of households, total population, and the household residents figures are available for all proposed Data Zones and Intermediate Zones. All census data outputs have privacy protection applied (known as cell key perturbation) to keep the data of individuals safe. In the case of 'nested' geography hierarchies, the impact of these Statistical Disclosure Controls (SDC) means the

sum of the population within Output Areas does not always match the corresponding Data Zones population totals.

Similarly, the sum of the population within Data Zones might not match the corresponding Intermediate Zones population totals. The same applies when summing totals at a local authority level, i.e. the sum of the populations of all Output Areas, Data Zones, and Intermediate Zones within a Local authority may not match the published Local authority total. Any new or reconfigured Data Zones or Intermediate Zones which are proposed via this consultation will be checked by the NRS Census team to ensure that the perturbed household residents totals meet the population thresholds. More information about the SDC applied to census data is available on the [Scotland's Census website](#).

## **Names and Codes**

We will contact local authorities to ask for names for Intermediate Zones (and local authorities can name Data Zones too, if they wish). In addition, official S-codes will be assigned using the 'S01' entity for Data Zones, starting at S01013482, and 'S02' entity for Intermediate Zones, starting at S02002515. Temporary codes have been used for the proposed Data Zones and Intermediate Zones in this consultation.

## **Summary of changes**

This section provides a summary of the proposed changes to Data Zones and Intermediate Zones. The proposed Data Zones and Intermediate Zones are available to view at [Proposed statistical geographies \(arcgis.com\)](#) and download at [Data Zones and Intermediate Zones consultation - Overview \(arcgis.com\)](#).

Currently there are 6,796 Data Zones and 1,279 Intermediate Zones in Scotland. The proposals covered by this consultation increase the number of Data Zones to 7,385, and the number of Intermediate Zones to 1,328.

There were 51 best-fit Data Zones which had fallen below 375 household residents and were merged with a neighbour. There were 317 best-fit Data Zones which were above 1,125 and were split into one or more Data Zones. There were 10 best-fit Intermediate Zones which had fallen below 2,135 household residents and were reconfigured. There were 51 best-fit Intermediate Zones which were above 6,375 household residents and were reconfigured.

If you require more detail on the changes made to proposed Data Zones and/or Intermediate Zones in your Local authority, please contact us at [statistics.enquiries@gov.scot](mailto:statistics.enquiries@gov.scot).

## **Consultation questions**

We are inviting respondents to propose changes to Data Zones and Intermediate Zones, as long as the criteria outlined above are met, i.e. proposals are within the population thresholds and are a continuous, compact shape. For Data Zones,

changes can only be made by moving entire Output Areas in or out of proposed Data Zones. This is because the Output Areas, which are combined to form Data Zones, are already published and therefore the boundaries are fixed. Intermediate Zones can be changed by amending the underlying Data Zones and/or moving entire Data Zones in or out of proposed Intermediate Zones.

- Are you content with the proposed 2022 Data Zones?
- Are you content with the proposed 2022 Intermediate Zones? Note that, if changes occur to the proposed 2022 Data Zones post consultation, these changes will be reflected in the Intermediate Zones.

## Responding to this consultation

We are inviting responses to this consultation by 1<sup>st</sup> October 2024.

Please respond to this consultation using the [Scottish Government's consultation hub](#). Access and respond to this consultation online. You can save and return to your responses while the consultation is still open. Please ensure that consultation responses are submitted before the closing date of 1<sup>st</sup> October 2024.

We intend to begin reviewing responses whilst the consultation is open, to ensure that the finalised boundaries can be published as quickly as possible once the consultation has closed.

If you wish to make suggested changes to the proposed Data Zones, please provide an explanation of the statistical benefits that would result from altering the boundaries. The most appropriate format for responses is a table, which could be in Excel format (a template is provided with the downloadable boundaries). The table should include the Output Area code, the proposed Data Zone code to which is has been assigned, and the code for the Data Zone into which you would like the Output Area to be moved (see example below), along with a short explanation of the rationale for the change. If you wish to create brand new Data Zones, please create new unique codes (for example, by adding a suffix to one of the proposed Data Zone Codes).

Example table requesting a change to move an Output Area

Output Area Code	Proposed Data Zone Code	Preferred Data Zone Code	Reason
S00136216	DZ330141	DZ330143	To align Data Zones more closely with community boundaries

Similarly, for Intermediate Zones, please provide a list of proposed 2022 Data Zones with their proposed Intermediate Zone assignments, and the codes for the

Intermediate Zones into which you would like the proposed Data Zones to be moved.

If you are unable to respond through CitizenSpace please contact us directly for assistance at [statistics.enquiries@gov.scot](mailto:statistics.enquiries@gov.scot) or:

Office of the Chief Statistician  
Scottish Government  
St Andrew's House  
Regent Road  
Edinburgh EH1 3DG

## **Handling your response**

If you respond using the consultation hub, you will be directed to the About You page before submitting your response. Please indicate how you wish your response to be handled and, in particular, whether you are content for your response to be published. If you ask for your response not to be published, we will regard it as confidential, and we will treat it accordingly.

All respondents should be aware that the Scottish Government is subject to the provisions of the Freedom of Information (Scotland) Act 2002 and would therefore have to consider any request made to it under the Act for information relating to responses made to this consultation exercise.

If you are unable to respond via the consultation hub, please complete and return the Respondent Information Form included in this document.

To find out how we handle your personal data, please see our privacy policy.

## **Next steps in the process**

Where respondents have given permission for their response to be made public, and after we have checked that they contain no potentially defamatory material, responses will be made available to the public on the consultation hub. If you use the consultation hub to respond, you will receive a copy of your response via email.

All responses will be analysed and considered along with any other available evidence to help us. Responses may be published where we have been given permission to do so.

## **Comments and complaints**

If you have any comments about how this consultation exercise has been conducted, please send them to the contact address above or at [statistics.enquiries@gov.scot](mailto:statistics.enquiries@gov.scot).

## **Scottish Government consultation process**

Consultation is an essential part of the policymaking process. It gives us the opportunity to consider your opinion and expertise on a proposed area of work.

You can find all our consultations online on the consultation hub. Each consultation details the issues under consideration, as well as a way for you to give us your views, either online, by email or by post.

Responses will be analysed and used as part of the decision-making process, along with a range of other available information and evidence. This will enable us to finalise the proposed changes to Data Zone and Intermediate Zone definitions.

While details of particular circumstances described in a response to a consultation exercise may usefully inform the policy process, consultation exercises cannot address individual concerns and comments, which should be directed to the relevant public body.



© Crown copyright 2024



This publication is licensed under the terms of the Open Government Licence v3.0 except where otherwise stated. To view this licence, visit [nationalarchives.gov.uk/doc/open-government-licence/version/3](https://nationalarchives.gov.uk/doc/open-government-licence/version/3) or write to the Information Policy Team, The National Archives, Kew, London TW9 4DU, or email: [psi@nationalarchives.gsi.gov.uk](mailto:psi@nationalarchives.gsi.gov.uk).

Where we have identified any third party copyright information you will need to obtain permission from the copyright holders concerned.

This publication is available at [www.gov.scot](http://www.gov.scot)

Any enquiries regarding this publication should be sent to us at

The Scottish Government  
St Andrew's House  
Edinburgh  
EH1 3DG

ISBN: 978-1-83601-525-3 (web only)

Published by The Scottish Government, July 2024

Produced for The Scottish Government by APS Group Scotland, 21 Tennant Street, Edinburgh EH6 5NA  
PPDAS1481158 (07/24)

W W W . g o v . s c o t