Statistical Spatial Framework

SSF Guidance Material – Using Geographic Boundaries and Classifications with Statistics

A typical statistic gives a numerical answer to a three-layered question:

- ♦ What was observed?
- ♦ Where was it observed ?
- ♦ When was it observed?

Location and geography are integral to answering the 'where' part of the question, but also must be associated with the when and the what.

Location

All statistics have a location associated with them. This can be very general such knowing that a statistic is associated with a country or a state. Alternatively, it can be more specific, such as a city or region, a suburb, a building or even a single point on the ground.

The location information associated with statistics are used for a range of purposes in the statistical process – collection, processing and validation, analysis and dissemination – with perhaps the most prominent being in the aggregation of data for external release (dissemination).

The Statistical Spatial Framework (SSF)[1] was developed by the ABS to provide a broad framework for integrating spatial (location and geographic) information with statistical information. It provides a distinct role for geographic boundaries and classifications by recommending the use of a common geography, the Australian Statistical Geography Standard (ASGS)[2].

Geographic classifications and boundaries

Geographic classifications and boundaries play a vital role in managing location information in the statistical process. A geographic classification breaks a very large area, such as a country, down into manageable pieces or regions. These classifications allow data to be summarised into regional statistics, so that the person using the data does not need to analyse all of the individual responses separately. This regional data can then be used to make statistical comparisons between regions or analyse trends across regions. Some geographic frameworks, such as the Australian Statistical Geography Standard (ASGS), have multiple levels or region types, which allow efficient and flexible release of data at levels that are suitable for various types of data and analysis needs.

---

[1] For more information follow this link: Statistical Spatial Framework (SSF)
[2] For more information follow this link: Australian Statistical Geography Standard (ASGS)

For statistical information, an ideal geographic classification must:

- ♦ Balance the size of regions so that detailed information can be released but also so that the privacy and confidentiality of the information can also be maintained.
- ♦ Be regions that are clearly identifiable or relatable to the data user (i.e. be known units, like suburbs, or be clearly mapped).
- ♦ Have well defined criteria for defining their coverage.
- ♦ Have boundaries that clearly define each region.
- ♦ Be suitable for releasing or analysing data across a range of statistical subject matter.
- ♦ Be comprehensive, covering the whole area to which the classification applies.
- ♦ Be stable over time to allow comparison of data over time.

The SSF recommends that data released or made available from any socio-economic dataset should include Australian Statistical Geography Standard (ASGS) regions. ASGS regions are the common geography in the SSF and will allow data from different sources to be integrated using these regions.

The ASGS is the ABS's official geographic classification framework. It provides the geographic classifications used in the majority of the statistical processes undertaken by the ABS, and the majority of socio-economic statistics released by the ABS include an ASGS region type. The ASGS also provides geographic classifications for use by National Statistical Services (NSS) partners and other organisations in their statistical datasets and data releases. Use of common geographic classifications across datasets enables data from different sources to be integrated at a geographic level. This allows data from different sources for a region or region type to be quickly compared and contrasted, or used as compatible inputs in more detailed analysis.

In addition to the ASGS, a range of regions that are not included in the ASGS are used by organisations in their statistical databases and in their statistical analysis (e.g. Medicare Local regions, school catchments, hospital regions, environmental zones, etc.). Each of these regions may or may not meet all of the criteria for an ideal geographic classification for statistics. Therefore, use of these geographies for statistical outputs should be considered carefully as it may limit the long term usefulness of the data produced. If data is released on geographies not included in the ASGS, the SSF recommends that data also be released for an appropriate ASGS region.

The ASGS

The Australian Statistical Geography Standard (ASGS) defines all of the standard regions used by the ABS to output data. This includes the main Statistical Area geography structure, which includes four sub-state region types, and a range of other ABS defined geographies based on urbanisation, remoteness, major metropolitan labour markets, and the Aboriginal and Torres Strait Islander population.

The ASGS also encompasses some key external boundary sets that the ABS releases data for, such as Local Government Areas, officially gazetted State Suburbs, Commonwealth and State Electoral Divisions and Natural Resource Management Regions.

Each level of the ASGS has been designed to maximise the amount of data that can be released for specific types of collections, while also minimising the risk that a particular person or organisation could be identified. Despite this, all externally released datasets must undergo rigorous confidentiality checks, including assessment of possible geographic differencing risks. For more information on confidentiality and geographic differencing, see SSF Guidance Material "Protecting Privacy for Geospatially Enabled Statistics: Geographic Differencing" on the NSS website.

ASGS replaces the ASGC

The ASGS replaced the Australian Standard Geographical Classification (ASGC), which the ABS used and maintained between 1984 and 2011. The ASGS was first released in July 2011 in time for the 2011 Census of Population and Housing (the Census). All other ABS data collections are migrating to the ASGS as their statistical cycles restart.

The main reason the ASGC was replaced was that it was unstable as a result of being tied to externally defined and constantly changing Local Government Area (LGA) boundaries for each state and territory. This meant that small area regions at the lower levels of the ASGC framework changed every year to follow LGA changes, which made it difficult and costly to change statistical processes and produce data that could be compared over time.

In contrast, the main structure of the ASGS will remain more stable, with changes made only due to population growth, changes in the built environment and for statistical classification reasons. Changes to boundaries will be made and published every 5 years with each Census. This will allow better comparability of data between geographic regions and over time.

The ASGS Structure

The ASGS comprises a hierarchy of geographic regions and it is split into two broads groups:
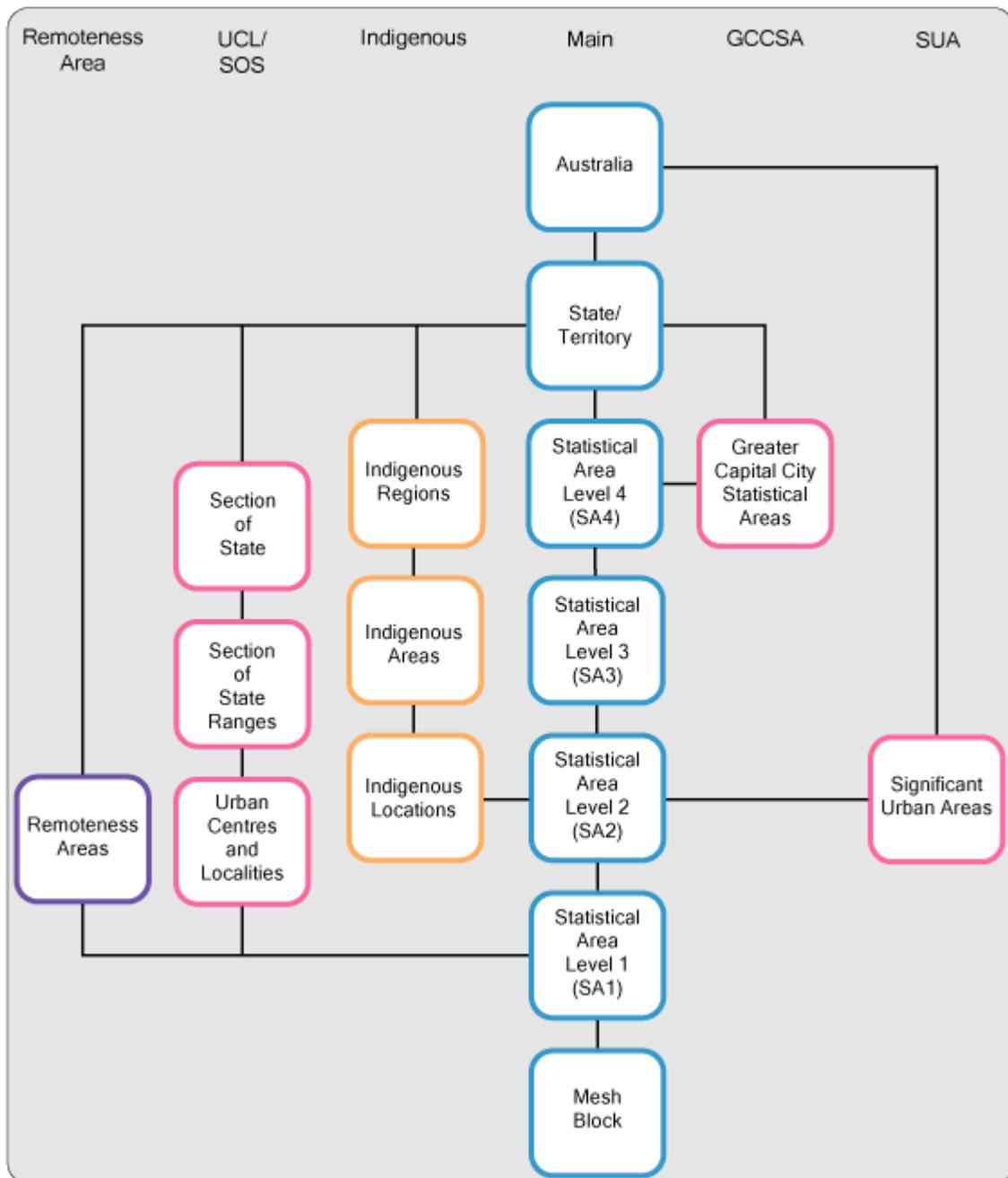
- ♦ ABS Structures
- ♦ Non-ABS Structures

ABS Structures

The ABS Structures comprise six interrelated hierarchies of regions. They are:

- ♦ Main Structure
- ♦ Indigenous Structure
- ♦ Urban Centres and Localities/Section of State Structure
- ♦ Remoteness Area Structure
- ♦ Greater Capital City Statistical Area (GCCSA) Structure
- ♦ Significant Urban Area Structure

ASGS structure diagram – ABS structures

Main Structure – ABS Structures

The ABS Statistical Areas in the Main Structure were designed to reflect functional rather than administrative areas. A functional area is the area from which people come to access services at a particular centre. The Statistical Areas have been designed to maximise the amount of detail available, while also managing statistical and confidentiality issues.

Mesh Blocks (MBs) are the smallest unit in the ASGS, and act as a stable, fundamental building block for all other ASGS regions. It is ABS policy that ABS data be coded to MB, wherever possible, and then released at aggregated levels from SA1 upwards. Mesh Blocks are not intended as a standard geographical unit for the release of data.

Statistical Areas Level 1 (SA1s) have been designed for the release of data for detailed small areas, such as Census data, and have a population range of 200-800 persons. SA1s differentiate between rural and urban landscapes and keep similar kinds of development together. Where possible, SA1s align to the officially gazetted suburb and locality boundaries.

Statistical Areas Level 2 (SA2s) have been designed as a general-purpose, medium sized area that represents a community that interacts together socially and economically. They have a population range of 3,000 – 25,000 persons, which is ideal for the release of demographic data such as the Estimated Resident Population. In major cities, SA2s reflect whole suburbs, groups of smaller suburbs or break up large suburbs.

Statistical Areas Level 3 (SA3s) provide a standardised regional breakup of Australia, they represent functional areas of towns over 20,000 people. They have a population range of 30,000 – 130,000 persons. SA3s were designed to create a standard framework for the analysis of ABS data at the regional level.

Statistical Areas Level 4 (SA4s) is the largest region before states and territories in the Main Structure and provides the best sub-state socio-economic breakdown in the ASGS. They have been designed for the release of Labour Force Survey statistics in particular, and have an optimum population range of 250,000-500,000 persons.

All other regions within the ASGS's ABS structures are aggregates of these areas (SA1s, SA2s, or SA4s), allowing for data integration across multiple region types without having to use mathematical transformations (correspondences – see below), which are not as accurate.

 For more information about the ASGS Main Structure, please refer to the ABS publication: 1270.0.55.001 - Australian Statistical Geography Standard (ASGS): Volume 1 - Main Structure and Greater Capital City Statistical Areas, July 2011

Other ABS Structures in the ASGS

The Other ABS Structures in the ASGS have been designed for specific purposes, the following table summarises each of these structures:

| Other ABS Structure	Purpose	Link (pages on the ABS website)
Indigenous Structure	Publication of statistics about the Aboriginal and Torres Strait Islander population of Australia.	Indigenous Structure Fact Sheet (PDF)
Urban Centres and Localities/Section of State Structure	Describes population centres with populations exceeding 200 persons	Urban Centres and Localities Fact Sheet (PDF)
Remoteness Structure	Divides Australia into broad geographic regions that share common characteristics of remoteness.	Remoteness Structure Fact Sheet (PDF)
Greater Capital City Statistical Area (GCCSA) Structure	Represent the functional extent of each of the eight state and territory capital cities.	Greater Capital City Statistical Area Structure Fact Sheet (PDF)
Significant Urban Areas Structure	Describes extended urban concentrations for more than 10,000 people which are clusters of related Urban Centres	Significant Urban Areas Structure Fact Sheet (PDF)

Non-ABS Structures in the ASGS

The Non-ABS Structures in the ASGS comprise eight hierarchies of regions which are not defined or maintained by the ABS, but for which the ABS is committed to providing a range of statistics. They generally represent administrative regions and are approximated by Mesh Blocks, SA1s or SA2s. They are:

♦   Local Government Areas
♦   Postal Areas
♦   State Suburbs
♦   Commonwealth Electoral Divisions
♦   State Electoral Divisions
♦   Australian Drainage Divisions
♦   Natural Resource Management Regions
♦   Tourism Regions

For fact sheets about Non-ABS Structures, follow this link: ASGS Fact Sheets

ASGS structure diagram – Non-ABS structures
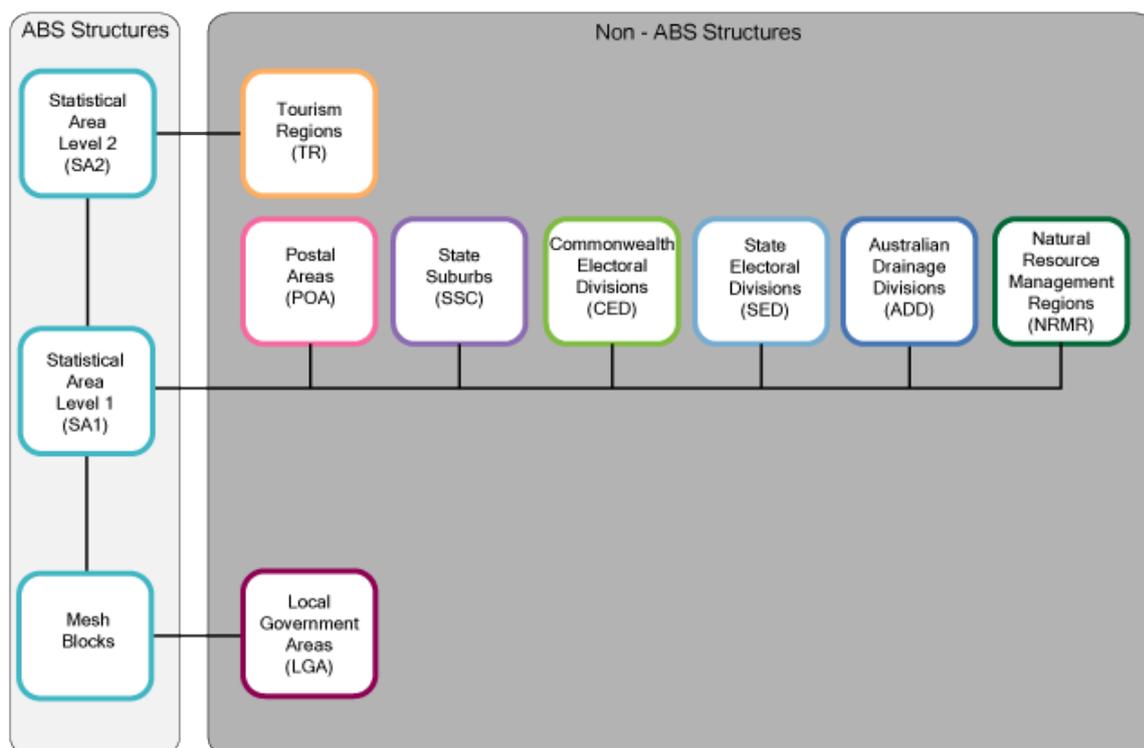
Selecting the right geography

Whilst the ASGS has been designed to be flexible for the release of different types of statistics, serious consideration needs to be given to what the most appropriate ASGS region is for different types of data. To select the most appropriate ASGS region for your data it is important to take into consideration the following issues early in the process:

- ♦ User needs
- ♦ Data quality – sample and non-sampling error
- ♦ Confidentiality
- ♦ Purpose of the data – current and future
- ♦ Cost
- ♦ Timeliness

The size and shape of a region impacts upon the picture the resulting data portrays. This can have a direct impact on the results of any subsequent analysis. The following link to an academic journal paper provides more information on this issue - Modifiable areal unit problem[3]

> The ABS Geography Section can advise on which ASGS geography is the most appropriate for your needs, please email: geography@abs.gov.au

---

[3] Openshaw, Stan (1983). "The modifiable areal unit problem", Norwick: Geo Books.
Alternative source for access - http://www.med.upenn.edu/beat/docs/Openshaw1983.pdf

Adding geographic information to your data

The Statistical Spatial Framework (SSF) recommends that each unit record in socio-economic datasets be geocoded with a location coordinate (i.e. latitude and longitude) and an Australian Statistical Geography Standard (ASGS) Mesh Block code. This geocode information is usually obtained through geocoding address information for each statistical unit in the dataset. The SSF also recommends that regional data released or made available from any socio-economic dataset should include ASGS regions – the common geography in the SSF.

The most appropriate method to add geocodes, such as ASGS regions, to your data depends on whether it is unit record level data or aggregated data and what geographic information the data already contains. The table below summarises different data types, the geographic information they contain and the appropriate transformation process that should be used to add geocode information. A diagram that shows the different pathways can also be found in the Appendix – Pathways for location and regional information.

Transforming geographic information across types

Data type	Geographic information contained in the data	Appropriate transformation process to be used	Geography available after transformation	Level of Accuracy
Unit record data	Location coordinate - Latitude and Longitude	Point-in-polygon allocation	Any	Most accurate
Unit record data	Building/site address	Geocoding	Any, via coordinate	Most accurate
Unit Record data	Locality information (e.g. Suburb and Postcode/State)	Coding Index	SA2 and above	Moderately accurate
Aggregated data	ASGS region	Allocation table	Higher level ASGS region	Most accurate
Aggregated data	Region information - other non-ASGS regions (e.g. Medicare Local regions, school catchments, hospital regions, environmental zones, etc.)	Correspondence	Higher or similar level ASGS region	Least accurate
Aggregated data	ASGS region	Correspondence	Region information other than ASGS	Least accurate

Unit record data

The options available for geocoding unit record data to the ASGS and other region types depends directly on the location information contained in the unit records.

For datasets that already have location coordinates attached to the unit records, these point references can be transformed into ASGS or other region types. This transformation uses GIS based point-in-polygon processes and allows data to be allocated to the selected region. Ideally, the point-in-polygon allocation process would assign an ASGS Mesh Block to each record, this then allows all other ASGS geography to be built-up from this basic building block using allocation tables.

Generally, the point-in-polygon allocation process is very accurate; however, the accuracy of the allocations is dependent on the accuracy of the original geocoding process used to obtain the location coordinates for each unit record. Therefore, it is critical to understand the accuracy of the coordinates included in any dataset.

Where a dataset has a full building or site address for each unit record then address geocoding is the most accurate method to obtain geocodes from the address information. Address geocoding will provide a location coordinate for each address and, ideally, an ASGS Mesh Block code. Region information can then be obtained using allocation tables and point-in-polygon processes. Further information about geocoding is contained in the SSF guidance material paper, "Geocoding Unit Record Data Using Address and Location" on the NSS website.

If only partial address information is available (such as suburb, state and/or postcode) coding indexes may be used to code unit records to ASGS SA2 units or higher level geographies. Locality, Postcode and State are all part of an address and when used in conjunction can effectively code unit record data to the SA2 level and above. This can be done using a suburb/locality to SA2 coding index.

There are several issues with using Postcode references alone to code data to the ASGS or other regions. There is no official geographic definition of Postcodes and they do not cover the whole of Australia. In general, Postcodes are larger than suburbs and consequently cannot effectively code data to the more detailed structures of the ASGS (e.g. SA1, SA2) and other smaller regions. However, it is possible to reasonably accurately code data to the larger SA4 and GCCSA levels, using a Postcode to SA4 index.

> Coding indexes are available from ABS Geography Section by emailing:
> geography@abs.gov.au

Aggregated data

Aggregated data is unit record data that has been summed together according to a characteristic or characteristics in the data, for example data for a region or for a particular age group. This data is commonly thought of in terms of data tables or reports, but can be presented in maps, graphs and other forms.

The options for converting aggregated data from an existing region to a new region depends on the region information the dataset already contains. If it contains:

♦ ASGS regions that need to be converted to a higher level ASGS region – use an allocation table.
♦ Other region information (not in the ASGS) that needs to be converted to an ASGS region – use a correspondence.
♦ ASGS region to be converted to a new region not in the ASGS – use a correspondence.

If the data already contains ASGS region information then allocation tables can be used. This is very accurate as smaller regions of the ASGS directly aggregate to larger regions of the ASGS that are in the same hierarchy. However, these tables can only be used to aggregate data to higher levels of the ASGS. They cannot be used to disaggregate data from higher levels of the ASGS.

If the dataset already contains geographic region information and you want to convert it to a new unrelated region then a correspondence may be used. Correspondences are a mathematical method of converting data from one region to another.

Correspondences are less reliable than address coding or allocation tables and the results can be misleading in some circumstances. This is because they are based on the assumption that the data to be converted is distributed evenly across the original regions or in line with distribution of the total population (see details below). Either of these assumptions may or may not be reasonable depending on the circumstances and the type of data being converted. Therefore, correspondences need to be used with a great deal of care. A quality indicator is incorporated in ABS correspondences to assist users in determining whether they are fit for purpose.

When correspondences are created the ABS uses a weighting unit to assist with the allocation of "From" data to their respective "To" units. The unit that is used to weight the correspondence will influence how the "From" units are apportioned, which in turn can have a major impact on the data values and hence the quality of the converted, or corresponded, data.

Population weighted correspondences are usually more effective for social and demographic data because populations are generally not evenly spread across a region. Area weighted correspondences are usually more suitable for agricultural and environmental data, because this data tends to be more evenly distributed across a region.

Correspondences are more accurate where the regions in the original dataset are smaller than the new regions the data is being converted to.

> For more information about correspondences and allocation tables, and to download files to convert data between regions, see the Correspondences page on the ABS website.

An example of a simple correspondence

This diagram is an example of a simple area based correspondence.

Region 1		Region 2	
A	*B*	*B*	*C*
100%	*61%*	*39%*	*100%*

The "From" regions are A, B and C, for which there is data. The "To" areas are Region 1 and Region 2, which are the areas that data needs to be converted to.

- All of the area of A is contained within Region 1, so 100% of its data is allocated to Region 1.
- 61% of B is contained within Region 1, so 61% of its data is allocated to Region 1.
- The remaining part of B is contained within Region 2, so 39% of its data is allocated to Region 2.
- All of the area C is contained with Region 2, so 100% of its data is allocated to Region 2.

If this is applied to the following data:

From Area	Persons
A	55
B	106
C	157

The data convertion is shown in the table below:

Persons	From	To	%	Converted
55	A	Region 1	100%	55
106	B	Region 1	61%	64.66
106	B	Region 2	39%	41.34
157	C	Region 2	100%	157

To produce the following results:

Region	Persons From A	Persons From B	Persons From C	Total Persons
Region 1	55	64.66	0	119.66
Region 2	0	41.34	157	198.34

Glossary

Please refer to the following link for a glossary of geographic and related terms used in this paper - LINK.

Where can I get further information?

More information on the Statistical Spatial Framework can be found on the SSF webpage on the NSS website – www.nss.gov.au

More information about geographic boundaries and classifications can be found by visiting the ABS website – www.abs.gov.au/Geography

Information about privacy and geospatial information is contained in the SSF guidance material paper, "Protecting Privacy for Geospatially Enabled Statistics: Geographic Differencing" on the NSS website.

Information about confidentiality is contained in the Confidentiality Information Series on the NSS website – www.nss.gov.au

Any questions or comments on this paper or other statistical geography topics can be emailed to geography@abs.gov.au

Appendix - Location and regional information pathways