ANZLIC
the Spatial Information Council

December 2020

# Guidance on the collection and use of data for location-based analysis

## Guidance on the collection of data for location-based analysis

When collecting data from entities such as individuals and businesses, the best-practice approach for government agencies regarding the collection of location data attributes is to:

1. **Collect a complete address** or the most comprehensive location data possible. This includes collecting address information as per the Australian Standard, *AS 4590.1: 2017 Interchange of client information*, of which the most common elements are:
    - Unit type and number
    - Level type and number
    - Address site name (may include building name or complex name)
    - Road number (can include textual prefixes and suffixes)
    - Street/road name (or water feature or island name)
    - Street/road type
    - Locality/suburb name
    - State/territory name
    - Australia Post postcode

    The collection of postcodes only can create issues when they are used for data analysis and is strongly discouraged (e.g. postcodes can change overtime, some postcodes cross state and territory borders and postcodes can vary in terms of population and geographical size). To enable meaningful location-based analysis of business intelligence data, the address data should be a physical location, not a Post Office Box address.
2. **Validate the collected location information (i.e. addresses) at the point of collection** using a trusted national source, such as the Geocoded National Address (G-NAF) in Australia, or the AIMS: Street Address Set in New Zealand. This will ensure the location information collected is accurate and reliable.
3. **Geocode the location information (i.e. addresses)** with a high degree of accuracy and consistency. This will maximise the value of the data collected, so it can reliably support location-based analysis of business intelligence data. For more information on geocoding, refer to the Australian Bureau of Statistics' (ABS) guidance on Geocoding Unit Record Data Using Address and Location.

## Address Validation and Geocoding Maturity Matrix

The table below describes the varying levels of maturity for the collection and use of data for location-based analysis.

| | Location input | Validation Process | Validation Data Source | Location Output |
|---|---|---|---|---|
| Optimal | Complete address | Immediate and automated address validation and assignment of geocoded information (e.g. latitude/longitude) | Authoritative address validation sources:<br>• Geocoded National Address File (G-NAF)<br>• Australia Post Address File (PAF) | Spatially accurate and high quality address and other geocoded information |
| Sufficient | Suburb/town/locality, state and postcode | | | Suburb/town/locality, state and postcode and likely administrative boundaries |
| Insufficient | Postcode only | Retroactive address validation and assignment of geocoded information | Bespoke address and location validation sources | Postcode and loosely approximated location information |
| Absent | No address or location | No address validation or geocoded information | No validation data used | No location information |

# Guidance on mapping and visualising data for location-based analysis

Mapping and visualising location-based business intelligence data using geospatial layers can enable comparison of different areas or regions. There are a range of geospatial layers available including the statistical areas defined by the Australian Statistical Geography Standard (ASGS) (e.g. mesh blocks, SA1, SA2, SA3, SA4 and states and territories), local government areas, localities/suburbs, electorates, or postcodes. New Zealand Statistical Areas as well as boundaries for local government areas and electorates are available from the Stats NZ Geographic Data Service.

There is no single 'one size fits all' solution when choosing a geospatial layer for mapping and visualisation. Each geospatial layer has advantages and disadvantages and users should assess these carefully in conjunction with the objectives and audience for their visualisation to determine the best approach. If integration of geospatial and socio-economic information is required, refer to the ABS' Using Geographic Boundaries and Classifications with Statistics. If the business intelligence data is sensitive, refer to the ABS' Protecting Privacy for Geospatially Enabled Statistics: Geographic Differencing.

The table below describes the different geospatial layers (from the most statistically accurate to the least) and their respective advantages and disadvantages. If a full address is collected, validated, and geocoded to a high level of accuracy, all geospatial layers will be available for analysis.

| Geospatial Layer | Description | Advantages | Disadvantages |
|---|---|---|---|
| **Statistical Areas (mesh blocks SA1, SA2, SA3, SA4, state and territories)** | Descriptions of the ASGS statistical areas are available on the ABS website. The ASGS statistical areas include:<br>• Mesh blocks – The smallest ABS-defined geographical areas, which build all other ASGS structures.<br>• SA1 – Areas with a population between 200 to 800 persons (average 400 persons), which aim to separate areas with different geographic characteristics within suburb and locality boundaries.<br>• SA2 – Areas with a population between 3,000 to 25,000 persons (average 10,000 persons), designed to represent a community that interacts together socially and economically (e.g. a set of related suburbs).<br>• SA3 – Areas with a population between 30,000 and 130,000 persons. SA3s report on regional data and are often the functional areas of regional cities and towns with populations greater than 20,000, or clusters of related suburbs around urban commercial and transport hubs in major urban areas.<br>• SA4 – Areas with a population above 100,000 persons, designed for reporting Labour Force Survey data.<br>• States and territories – Represent all of Australia, divided up into each of its states and territories. 'Other Territories' includes Jervis Bay Territory, the Territories of Christmas Island, Cocos (Keeling) Islands and Norfolk Island. | • Provide an integrated set of standard areas for analysing, visualising and integrating data and statistics, including those the ABS and other organisations produce.<br>• Units of each statistical area are readily comparable across time and are statistically robust, for time series comparisons and visualisations.<br>• Users can transform data (and associated maps/visualisations) from one ASGS statistical area to others, using ASGS correspondences. This is particularly useful when data is only available at a geographic area level rather than at the geocoded address level. | • Statistical areas can be unfamiliar to the public and may not suit all audiences.<br>• Can require substantial statistical and geographic expertise to 'clean up' non-ASGS data and convert the data into ASGS statistical areas.<br>• Work well when data is available at the complete address or mesh block level, but does not work well for data that has few or no geospatial attributes or identifiers (e.g. only state/territory, LGA and/or postcode level data).<br>• When transforming data, users should transform data from smaller statistical areas to larger ones. Doing the reverse will introduce a high level of inaccuracy to the data. |
| **Localities/ Suburbs** | Localities are the officially recognised boundaries of suburbs (in cities and larger towns) and localities (outside cities and larger towns). Gazetted Localities data is openly available as part of the Administrative Boundaries dataset on data.gov.au. The Fire and Emergency NZ Localities dataset provides suburbs (urban areas) or localities (rural areas) for all of New Zealand and is available under Creative Commons licensing via the LINZ Data Service. | • Well-known and recognised by most individuals/entities, and as a result are easy to collect.<br>• Often cover relatively small areas or regions, making them well suited to visualising data or analysing factors in localised areas or regions. | • Locality boundaries can change over time and may be difficult to trace over time.<br>• The definition of a locality differs between states and territories, e.g. in New South Wales, locality corresponds to suburb whereas in Queensland it is defined as both a bounded area, considered being of a rural or semi-rural nature within a local government area, and a suburb – a bounded area within a City, Town, Shire, or Region, principally of urban character.<br>• There are over 2,000 locality names duplicated across Australia. Many of these duplicates exist within the same state or territory boundary. |
| **Local Government Areas (LGAs)** | Represent gazetted local government boundaries within each Australian state and territory, as defined by each state and territory local government department.<br>Given local governments define LGAs based on their areas of service delivery responsibility, LGAs can change somewhat arbitrarily over time and are not well suited to time series comparisons within or between LGAs. | • Generally recognised by the public / non-specialised audience.<br>• Supports visualisation of geospatial data to provide an understanding of factors in a given area at a point in time. | • Time series comparisons can be complex, given LGAs can change arbitrarily over time (e.g. via local elections or changes to areas of responsibility for local governments).<br>• Some localities or suburbs can cross over two or more LGAs.<br>• Poorly suited to comparative analysis, given some LGAs are very large and sparsely or unevenly populated, while others are very small and densely populated.<br>• LGAs are not always contiguous. |
| **Electorates** | An Australian Electoral Commission (AEC) electoral division boundary is an area legally prescribed for returning one member to the House of Representatives, Australia's Federal Lower House of Parliament. Further information is available on the Australian Electoral Commission website: http://www.aec.gov.au.<br>The states and territories also have electoral commissions that produce their own electoral districts. The ABS' ASGS Commonwealth and State Electoral Divisions approximate these districts using one or more Statistical Areas Level 1s (SA1). | • Generally recognised by the public / non-specialised audience.<br>• Well known by ministerial and political audiences, given electorate boundaries use for electoral polls. | • Electorates can change between each election, as determined by the electoral commissions.<br>• Electorates are not designed for statistical comparisons.<br>• Electorates are best suited to political audiences.<br>• Some AEC electorates cross state and territory boundaries (e.g. Norfolk Island is included in the ACT electorate of Bean; and Christmas Island and Cocos (Keeling) Islands are included in the NT electorate of Lingiari). ASGS State Electoral Divisions do not cross state and territory borders. |
| **Postcodes** | A postcode is a four-digit number used by Australia Post to assist with mail delivery. | • Well-known and recognised by most individuals/entities, and as a result can be easily collected.<br>• Provide an indication of mail delivery areas, which can be useful to inform localised survey activities. | • Some areas do not have a postcode assigned to them.<br>• Postcodes can change over time. This makes postcodes a poor choice for visualising time series data.<br>• Some postcodes cross state and territory borders.<br>• Postcodes are highly variable in terms of how many people live in them and their geographical size.<br>• Some postcodes are not street delivery areas. These include post office boxes, mail back competitions, large volume receivers and specialist delivery postcodes. These postcodes are only valid for postal addresses and are not a valid location for population data.<br>• Postcode areas are not always contiguous.<br>• Defining postcodes with a geographic boundary is an imprecise process, as they are a set of delivery units and not designed as spatial units. This results in substantial variations in postcode boundaries that different organisations release. |