Data Linkage Branch

November 2018

# Reference material: DLB geocoded data items

## Overview

The Data Linkage Branch (DLB) uses a spatial software package called Spectrum to perform geocoding. Spectrum is a product created by Pitney Bowes.

This process involves the allocation of latitude and longitude to records on the basis of residential address. Geographical boundaries (SA1, SA2, CD, SLA, LGA) and associated indices (SEIFA, RA) are subsequently generated for each applicable census year.

## Limitations

There are a number of unavoidable limitations that should be observed when performing spatial analysis using the DLB geocoding fields:

* The geographical boundaries (SA1, SA2, CD, SLA, LGA) are sometimes realigned from one census to the next. For this reason, direct comparison between different census years is not recommended. Although all census/boundary combinations are provided as a courtesy, DLB recommends using the census year of closest proximity to the date of the record.
* Records that yield a poor quality spatial match will be assigned a region (e.g. an SA1) based on the centroid of a potentially large geographical area. This area may not overlap entirely with the assigned region. Therefore, some regions may be assigned in error.
* In some instances, an address may be matched to a location that straddles a border between boundaries. In this instance, either boundary may be assigned.
* The geocoding process is dependent upon the quality of the address data, and in some cases an address may be difficult or impossible to geocode accurately, or at all.

## Interpretation of MatchCodes and MatchScores

The Spectrum Geocoder outputs two different measurements of match strength depending on how the match was found. If the match is found through the general geocoding module it outputs a MatchCode. MatchCodes contain a series of letters and numbers to denote how close the match was to a real address.

For example, the best possible MatchCode is 'S8HPNTSCZG'

This means that the address matched on;

* single parcel level (S8)
* house number (H)
* street prefix (P)
* street name (N)
* street type (T)
* street suffix (S)
* city or suburb (C)
* postcode (Z)
* GNAF12 address (G).

A minus sign '-' in any of the above values denotes that there was no match on that level. There are different variants on the above codes, to demonstrate other extents of matches.

More information on the different codes can be found at the Pitney Bowes website. The accompanying document (Spectrum\_11.0\_Geocoding\_AUS.pdf) describes the codes in detail in Section 3.

## Aboriginal Communities and Named Establishments

**Please note: DLB is currently revising some of its geocoding process flows. As of 21 November 2018, the matching process described below is not being applied, and therefore MatchScores are not being assigned.**

If the match is found through the value adding Aboriginal Community or Establishment geocoding module (which geocodes Aboriginal Communities or Establishments) then Spectrum outputs a MatchScore. The MatchScore is a percentage value that shows how well the text of the Aboriginal Community or Establishment, matched to that of a known Community or Establishment.

## Further information

Further information about Spectrum can be found at the Pitney Bowes [website](https://www.pitneybowes.com/au/location-intelligence/geocoding-data-enrichment/spectrum-enterprise-geocoding-module.html).

The boundaries, SEIFAs and RAs are developed by the Australian Bureau of Statistics (ABS), and the Data Linkage Branch uses mapping files available through the ABS website to attach them to the data. Queries about their interpretation and use should be directed to the ABS. For further information, please visit the ABS website at [www.abs.gov.au](http://www.abs.gov.au).