

Geographic Information in IDI

Note: IDI changes often. The information in this document was current as at December 2017.

Introduction

Information about where individuals live is important for many analyses. For example we may want to examine health outcomes by region (e.g. DHB) or by level of deprivation using NZDep. Sometimes we want to use address information more directly, for example in constructing estimates of residential mobility (the number of addresses an individual has lived in over a given time-frame).

What geographic information is available in IDI?

IDI contains information about the address updates that individuals have provided to various government agencies, and addresses recorded in surveys (such as census). This information is geocoded (text addresses are converted to standard geographic locations) by IDI as part of the refresh process. Geocoded address information is available for several geographic classifications:

- Address uid (e.g. `snz_idi_address_register_uid`): This is an encrypted version of NZ Post's National Postal Address Database identifier (essentially a postal address). A unique address is represented by the same encrypted identifier throughout the IDI.
- Meshblock (e.g. `ant_meshblock_code`): The smallest geographical area in NZ standard geographic classification, representing roughly 30 to 60 dwellings. Meshblocks are combined to create higher geographies such as area units, territorial authorities, DHBs, and regions. As areas grow and change, meshblocks get updated (for example, a rapidly growing meshblock may be split into two smaller meshblocks). The annual update of meshblocks is referred to as a "meshblock pattern" and is usually referenced by date eg `mb13` (the 2013 update), `mb17` (the 2017 update). Meshblocks in IDI are coded to the most recent pattern available at the time of the refresh.
- Territorial authority (eg `ant_ta_code`): A larger area defined as a city council or district council. There are 67 territorial authorities consisting of 12 city councils, 53 districts, Auckland Council, and Chatham Islands Council.
- Post code: Post codes are available in some datasets in IDI. Postcodes do not map exactly to standard geographic classification (you cannot combine meshblocks to create a postcode), and can cover very large geographic areas (eg in rural delivery areas) and therefore are not recommended for geographic analyses. Consider using territorial authorities or regions instead.

Additional geographic classifications

If you want to use additional geographic classifications such as DHB or ward, or area-based indexes such as NZDep, you will need to match the geographic information in IDI against other concordances.

The meshblock_concordance table on the IDI Metadata server contains a concordance of annual meshblock patterns, allowing you to convert (for example) mb17 to mb13. This is useful when using NZDep, which is only produced for mb06 and mb13. (A NZDep concordance is also stored on the metadata server. For more on NZDep [see here](#)). Full concordances for DHB, data zones (see below), and other classifications are not currently available on any of the main IDI servers. Where concordances are not available within the IDI, you will need to obtain the concordance files yourself and have them transferred to your datalab folder (by emailing access2microdata@stats.govt.nz). Area concordances can be downloaded from the [Stats NZ website](#).

Dan Exeter and colleagues have recently developed an area classification called 'data zones'. In general data zones are larger than meshblocks but smaller than area units. They are designed to be used with the Index of Multiple Deprivation, a measure of socioeconomic deprivation. Concordances for converting meshblock to data zone are not currently available in the IDI, but can be downloaded [here](#).

Address notification tables

Address information from Ministry of Health (PHO and NHI registers), Ministry of Social Development, Ministry of Education, ACC, and Inland Revenue is collated into two 'address notification' tables:

- address_notification_full is a collation of all address updates notified to the above data providers. The most useful variables in this table are 'ant_notification_date', which is the date that the address was updated with the data provider, and the geographic variables: address uid, meshblock, territorial authority, and region.
- address_notification is a prioritised version of the address_notification_full table. To create this table, address sources are split into two quality tiers based on characteristics of the source data (such as whether the agency intends to record residential addresses, how often they update addresses). Addresses from the higher quality tier are prioritised. In general, once an individual has an address recorded from the high-quality tier, new updates from the lower-quality tier will not be added to address_notification: the address can only be replaced by an update from a high-quality source. (See metadata on the IDI wiki for more details).

In practice there is not much difference between these two tables: most address updates come from the high quality sources, so most updates are included in the prioritised address_notification table. The address_notification_full table allows for the greatest flexibility as it contains all address updates, allowing researchers to select the updates that they want to use. The code on the VHIN website uses the address_notification_full table.

In addition to the address notification tables, IDI contains other sources of geographic information such as census night address, and addresses from other surveys. This information is stored in the relevant tables (eg census or SoFIE), rather than in the address notification tables. Typically the addresses recorded in surveys relate to a specific date. Depending on your project requirements you may want to consider using them.

How can I find out where someone lives at a given date?

The VHIN website contains a piece of code that determines the most recently updated address at a given date ([see Code Sharing guide](#)). This method uses all address sources from the `address_notification_full` table. If an address is not recorded prior to the reference date, the first address update within the 12 months following the reference date is used.

The above method is used in a range of projects and gives a good approximation of address at a given date. However, depending on the specific requirements for your project you may wish to vary the above method by (for example) excluding some address sources, or excluding address updates that occur after the reference date.

How can I measure residential mobility?

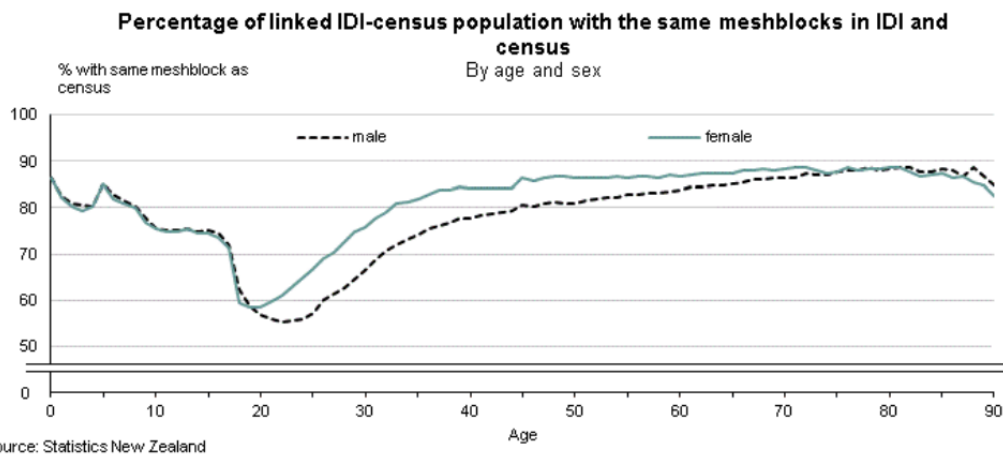
Sometimes we want to know the number of times an individual has moved residence in a given time period. We call this 'residential mobility'. In theory it is possible to measure residential mobility by summing the number of updates in the address notification table over a given time period. In practice, this produces large numbers of address changes for some individuals, and we do not know if these are genuine. A better alternative may be to count up the number of distinct `address_uids` that an individual has lived in over a given period.

Limitations of address information

The address data in IDI has several limitations when we are trying to use it to establish where an individual is living:

- Addresses given to government agencies may not represent an individual's place of residence. For example, individuals may provide a postal address where they want their mail delivered, or another proxy address for correspondence. This varies between agencies: some agencies aim to collect residential addresses, others require a contact address only.
- Some individuals (e.g. children in shared custody arrangements) may have more than one usual residence. Dual residences are difficult to identify in IDI and these individuals may appear to be constantly moving back and forth between residences.
- Most of the address records are notifications of address updates, and as such the dates attached are the date that the administrative provider was notified of the address change, not the actual date that an individual moved address.

Some information about the quality of IDI geographic information comes from a comparison of IDI addresses against Census addresses. The IDI meshblock information in the `address_notification_full` table was consistent with the census for 79% of people, suggesting reasonable quality. This was higher still for larger geographical areas. The level of accuracy differed by age and sex (See Figure), with lower consistency for the young adult group.



Source StatsNZ: Gibb (2015)

Summary

IDI contains a range of geographic information. This information can be used to identify meshblock of residence, DHB, area level deprivation, residential mobility and can also be used to link environmental or other area-based data to an IDI population. The address_notification_full table in the IDI is a helpful aggregation of geographic information, and it is reasonably consistent with the census address information. A method for determining an individual's address at a given point in time is available on the VHIN website. However, IDI address data has limitations and these should be kept in mind when using it for research.

Further reading

Gibb, S. J. & Das, S. (2015). *Quality of geographic information in the Integrated Data Infrastructure*. Retrieved from www.stats.govt.nz.
<http://archive.stats.govt.nz/methods/research-papers/topss/quality-geo-info-idi.aspx>

'Geospatial information in the IDI', available from IDI wiki.

Geographic definitions on Stats NZ website: <http://archive.stats.govt.nz/Census/about-2006-census/2006-census-definitions-questionnaires/definitions/geographic.aspx>

Geographic area files, downloadable from
http://www.stats.govt.nz/browse_for_stats/Maps_and_geography/Geographic_areas/geographic-area-files.aspx

Exeter DJ, Zhao J, Crengle S, Lee A, Browne M (2017) The New Zealand Indices of Multiple Deprivation (IMD): [A new suite of indicators for social and health research in Aotearoa, New Zealand](#). PLoS ONE 12(8): e0181260

Zhao J, Exeter DJ. Developing intermediate zones for analysing the social geography of Auckland, New Zealand. *New Zealand Geographer*. 2016;72(1):14-27

Original web post by Sheree Gibb, Andrea Teng 8/12/17