

# Geocoding and De-identified Data

## Geocoding

The geo-coding process is achieved by the following steps:

1. CAT separating the patients address from de-identified demographic and clinical information to ensure they are never held anywhere together outside the practice.
2. CAT tagging the de-identified data with a unique and encrypted code so on arrival in PAT CAT it can be matched to a geographical location. This code has no meaning outside the PAT CAT system.
3. CAT sending the address securely to an Australian based mapping service which returns the geospatial location information for each address for usage by PAT CAT.
4. These codes are used to create a map in PAT CAT utilising the geospatial location information matched with the de-identified data to create a map displaying the number of patients with a specific profile and how they are distributed. This provides the ability to correlate patient problems with location and look for patterns.
5. PAT CAT will not display patients in an area where the number drops below 5.

To ensure patient privacy, all mapping of patient data in PAT CAT is of aggregated data, individual patients are never shown. In fact it requires at least five patients to be present within a map segment (SA2) for any information to be shown.

## Data De-identification

The data de-identification and geocoding process occurs through the De-identify Dataset / Add GTAGs functionality included in CAT4. When an extract is run through the De-identify Dataset / Add GTAGs functionality;

- all patients who have withdrawn consent for their data to be used in clinical research are removed;
- the extract is de-identified by removing all identifiable information such as name, address and date of birth. The only personal information which remains in a de-identified extract is gender, ethnicity and the age of a patient in years.

de-identified data refers to patient clinical data (e.g. diagnosis, measurements, pathology etc) that does not contain information that can be used to reverse identify the patient. De-identified data is permissible for use where it is not practicable to obtain informed consent for use of data. PEN CS has provided tools within CAT4 that can be run by the practice to create a de-identified patient data set. These tools are collectively known as FAT CAT, the Filtering and Anonymisation Tool for CAT. They are incorporated into CAT4.

FAT CAT ensures that the de-identified data extract produced by Classic CAT strictly adheres to the National Privacy Principles.

- The Filtering process removes any patients who have withdrawn their consent to share data from the de-identified dataset. These patients can be flagged within CAT (Refer [CAT Patient Consent Withdrawn Quick Reference Guide](#)).
- The Anonymisation process de-identifies the data. All patient information that is identifiable or partially identifiable is removed from the de-identified extract (e.g. name, address, postcode, date of birth, Medicare number etc).

Classic CAT de-identified data files that have been through the FAT CAT process are tagged as "certified". PAT CAT recognises FAT CAT "certified" data files and these are the only data files that PAT CAT will import and display and rejects the files that were not FAT CAT "certified".