

# 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- all patients who have withdrawn consent for their data to be used in clinical research are removed;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- The Filtering process removes any patients who have withdrawn their consent to share data from the de-identified dataset. These patients can be flagged within CAT (Refer [CAT Patient Consent Withdrawn Quick Reference Guide](#)).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".