*Database Development*

This phase is focused on analysing and aggregating the collected data to be able to build the Charging Station Database that is the primary database used in the website.

1. Define Charging Station database fields and the valid values for each field
   1. Data Dictionary was created to avoid incorrect interpretation of data used in the website
   2. The Data Dictionary will also be the reference document in the Data Maintenance Phase
2. Open data from the government and data from not-for-profit organisation will be collected first, to be followed by commercial establishment related data
3. Conduct data exploration and analysis for each data source
   1. Determine the data source size and complexity to identify if the environment for data exploration/analysis
      1. MS Excel for small and straightforward data
      2. Python for medium to larger complex data
   2. If a data source is not a csv downloadable file, convert the data sources into a data table
   3. Identify fields available for each data source that is aligned with the charging Station data fields
4. Conduct data cleansing
   1. Determine filter rules for each field to be able to include relevant data into the Charging station database
   2. Validate if the locations data are correct using reverse geocoding and normal geocoding method
5. Fill missing data
   1. Retrieve missing location using reverse geocoding and normal geocoding methods
   2. Apply rules on populating missing location data
      1. The open data from the gov’t is always assumed correct for longitude, latitude and full address
      2. Geocoding results are always selected for postcode and suburb data as its data were more
6. Map the columns of the cleansed data into the Charging Station data table
7. Visualise the data to determine any data outliers/errors and fix them