Open Data Methodology -

Title of data set	Small Scales Generation		
Methodology Author	Ermelindo Enoque	Methodology data	02/04/2025
Brief Description of data set	This dataset provides a general overview of the integration of low carbon technologies across Northern Powergrid distribution licence areas. The data has been filtered to include only three fields due to sensitivities in the data which were identified during data triage. The data is produced in accordance with Ofgem guidance to comply with electrical distribution standards.		
Meta Key words	Low Carbon Technologies		
Relevant User and Theme pages	Connections team. New connection		
Scope of dataset	This dataset is compiled from third-party data received from different customers who applied for new connection and stored within Northern Powergrid connections team (QPID).		
Data set structure/granularity/field descriptions	This dataset is in excel format, with each row corresponding to a primary substation (GeoJSON format) in Northern Powergrid Distribution Licence areas, and each column represent a specific attribute. In terms of granularity of the Small Scales Generation data, the low carbon technologies notification has been aggregated by power capacity linked to each primary substation boundary area. **Primary substation** – is where the transformer voltage from extra high voltage (EHV) is dropped down to high voltage (HV), to ensure that the electricity is distributed efficiently across area cover/boundary area covered by the substation. **Number of notifications** – refers to the number of notifications that have been requested for new photovoltaic generation connections, enabling Northern Powergrid to track the adoption of low carbon technologies within our licence area. **Total capacity** – shows the cumulative addition of generation of all newly connected generators to each primary substation.		
Source of data in dataset	The dataset is received from Northern Powergrid's connection team every month via email. This data has various associated schemes and, for Small Scales Generation specifically we use the G98 and G99 schemes which is for photovoltaic (PV) generators.		

Update frequency and maintenance plan	A new dataset is received and update every month to meet Ofgem requirements/ Licence obligation.
Data Quality /completeness /accuracy	We are confident that for our low carbon technologies the data is of good quality. In terms of completeness, considering the three-field listed above, they are all present in the published dataset.
Related datasets	Photovoltaic generators

Methodology and assumptions

We load the LCT notification excel file and GeoJSON file of the primary substation boundaries to our Azure Databricks workspace, then filter by G98 & G99 schemes. We then utilise python libraries to extract latitude and longitude coordinates from postcode data and overlay the primary substation boundaries with the LCT notifications data. Following this, we then aggregate the power generation by each scheme, to a corresponding primary substation, filtering by the three columns table. We then connect Databricks to our Open Data Portal to publish the dataset.

Detailed considerations/exceptions/limitations

Inaccurate and inconsistent data entry, particularly in the primary substation column, can lead to gaps in the imported power capacity to each substation. This should be considered during planning purposes and impact assessments on the network/ grid.