# The Challenge

CATAPULT
**Energy Systems**

- Various organisations have made efforts to publish open datasets about the UK's energy system, e.g. National Grid ESO, Elexon, UK Power Networks, etc.

- Regardless, some fundamental work to join up datasets about assets in the energy system appears to be repeated between different solutions

- There is an active open-source data science community trying to build useful tools and insights with this data, for example:
  - Power Station Dictionary (PSD)
  - Wind Curtailment Monitor (WCM)

## Beatrice

### Identifiers

| Relationship | ID Type | ID(s) |
|---|---|---|
| Root | Dictionary ID | 10297 |
| Related | Settlement BMU ID | T_BEATO-1, T_BEATO-2, T_BEATO-3, T_BEATO-4 |
| Related | National Grid BMU ID | BEATO-1, BEATO-2, BEATO-3, BEATO-4 |
| Related | 4C-Offshore ID | united-kingdom/beatrice-demonstration-united-kingdom-uk46, united-kingdom/beatrice-united-kingdom-uk53 |
| Related | WindPowerNet ID | windfarm_en_10664_beatrice-demonstration, windfarm_en_12051_beatrice |
| Related | EIC ID | 48W00000BEATO-1T, 48W00000BEATO-2R, 48W00000BEATO-3P, 48W00000BEATO-4N |
| Related | CfD ID | INV-BEA-001, INV-BEA-002 |
| Equivalent | Common Name | Beatrice |
| Equivalent | Wikidata ID | Q4877211 |
| Equivalent | Wikipedia ID | Beatrice_Wind_Farm |
| Equivalent | Power-Technology ID | beatrice-offshore-wind-farm |
| Equivalent | REPD ID (Old) | A0275 |
| Equivalent | REPD ID (New) | 2521 |

# Our Objective

- Demonstrate the value (and challenges) of building upon existing open-source projects like PSD when developing a common energy system use-case
- Contribute to the open-source community by developing a useful new dataset that would allow us to map near-real-time generation by location across the UK
- Connect data scientists with an interest in similar datasets to highlight overlap and collaboration opportunities
- Contribute our findings to the existing open-source projects

# Our Approach

- ✓ Explore source code and contribute to the PSD
- ✓ Identify potential input datasets for the live-generation map
  - ✓ Generation asset IDs & locations
  - ✓ Data about generation from the "Balancing Mechanism Reporting Service" (BMRS / Elexon): historic and current
- ✓ Data design and confirmation of assumptions: reconciliation between historic and near-real-time data
- ✓ Development of the live dataset
- ✓ Reconciliation of different datasets with wind farm locations and generator IDs

In progress:
- Mapping of live data
- Publication of dataset code, documentation and maps

# Outcomes – Wind Farm Locations

- Identified multiple discrepancies between PSD and WCM wind farm datasets including:
  - Missing generators (in either dataset)
  - Incorrect data about generators (i.e. installed capacity or location)
- Made contributions to the PSD and WCM to address identified data quality issues

| Issue Type | PSD | WCM |
|---|---|---|
| Wind farm missing | 6 | 31 |
| Potential Location discrepancy >20km | 35 (TBC correct location based on further investigation) | |
| Installed capacity missing | - | 42 |
| Potential large discrepancy in installed capacity | 3 | 17 |

# Outcomes – Live Generation Data

- Reconciled historic generation data made available by Elexon ("B1610 - Actual Generation Output Per Generation Unit") with current Balancing Mechanism data ("Physical Data") to confirm approach to generating the "live" dataset.

- Reconciliation results across 7 days (January 24-30, 2023): 336 (7*48) 30-minute settlement periods (SP) with 251 generators

| % error | % of generators across entire period | % of Records (SP per generator) | % of overall generation |
|---|---|---|---|
| <=1% | 36% | 28% | 78% |
| 1-5% | 29% | 16% | 13% |
| 5-10% | 16% | 8% | 6% |
| >10% | 19% | 48% | 3% |

- 78% of generation came from conventional generators with high data accuracy (<1% error for the generator across the observation period): mostly predictable, conventional generation
- Some wind farms, especially smaller ones, fail to provide accurate forecasts to National Grid.
- Due to these inaccuracies, NG devises its own forecasts for wind farms which are more accurate than the BMRS data: future work for the community could focus on this aspect of the tool.

# Outcomes – Live Generation Map

- [Placeholder – add visualisation once initial version has been completed]

# Learnings

- Open-source mappings between different IDs are invaluable, but hard to maintain and publicise

- Data documentation that focuses on individual datasets is insufficient – most value comes from combining datasets and nuances around this are often not captured

- Well documented example open-source data pipelines that combine key datasets seem a good way to address this

- When building foundational open-source data tools it's important to provide documentation focused on end-users

- Skillsets required to carry out initial analysis and build prototype differ from skillsets required to productionise solutions, which makes it easy for open-source solutions to fail to deliver long-term value

- Ongoing support and maintenance of developed open-source solutions needs to be considered