Data for Future Energy Systems

Christopher Jackson d'Ing (INSA), CEO
Advanced Infrastructure

Our Team of 25+ energy and software engineers

CEO
Christopher Jackson

CTO
Suhas Dattatreya

CPO
Lily Cairns Haylor

COO
Jonathan Jackson

Senior Scientist
Dr Alastair Buckley

Senior Scientist
Dr Jamie Taylor

Senior Scientist
Dr Julian Briggs

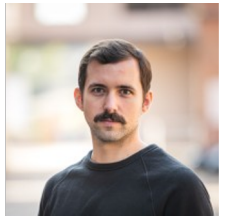
Senior Scientist
Dr Robert Brown

Energy Modeller
Dr Hindolo George Williams

Software Engineer
Kalpitha Nidigal

Energy Modeller
Dr Noah Miller

Advisor
Ashley Barrett,



Advisor
Dennis Birkhoelzer,

Advisor
Freerk Bisschop,

Our Partners

nationalGrid sse DB eex NGI SMA UK Power Networks Delivering your electricity

Our Challenge: The energy transition is falling behind

The EU needs to spend €400 billion euros per year on low carbon infrastructure but is only spending half of this target energy planning is a human-resource heavy industry with large data silos

100% low carbon heating by 2050

Less than 5% of homes are heated this way today. “At current rates, it would take 700 years.” We need to transition 1 million homes per year

3 million EV Chargers by 2030

Are needed, according Committee on Climate Change (CCC). Today, there are 330,000. We need to install 400,000 chargers per year

100% renewable generation by 2050

In 2022, 45.5% of generation is renewable. Electrification is increasing the challenge. We need 550-680 TWh of low carbon generation needed to meet demand in 2050

1. **Technical scope**
2. Selection of algorithms and tools
3. Scalability & flexibility
4. Data security and legal compliance
5. Quality assurance and Risk management
6. Degree of innovation of the technical solution
7. Status of the mock-up & Product Demo

Our solution is to build Enterprise SaaS tools for Power Utilities to help remove data barriers in the energy transition.

Our goal is to build a platform for exchange of cross-sector spatial data for the purpose of building retrofits to help speed up the energy transition.

The platform connects proprietary industrial data along with personal data across buildings, energy and transport to support decisions-making for faster deployments retrofits

1. 90% time saving spent sourcing data
2. 30% better engineering decisions
3. 1% saving on network reinforcement

1. **Technical scope**
2. Selection of algorithms and tools
3. Scalability & flexibility
4. Data security and legal compliance
5. Quality assurance and Risk management
6. Degree of innovation of the technical solution
7. Status of the mock-up & Product Demo

The service is a trusted and secure privacy-aware analytics solutions allowing for the secure sharing of proprietary industrial data along with personal data between Power Utitiies and other stakeholders.

This innovation supports Secure Data Value Chains and aligns with the outputs of H2020 PLATOON (Digital PLAtform and analytic TOOls for eNergy)

The diagram shows how the solution maps against the PLATOON Reference Architecture. Outputs will be compatible with PLATOON Data Analytics and Processing tools, Data Connectors & Ingestion functions in the Interoperability layer.

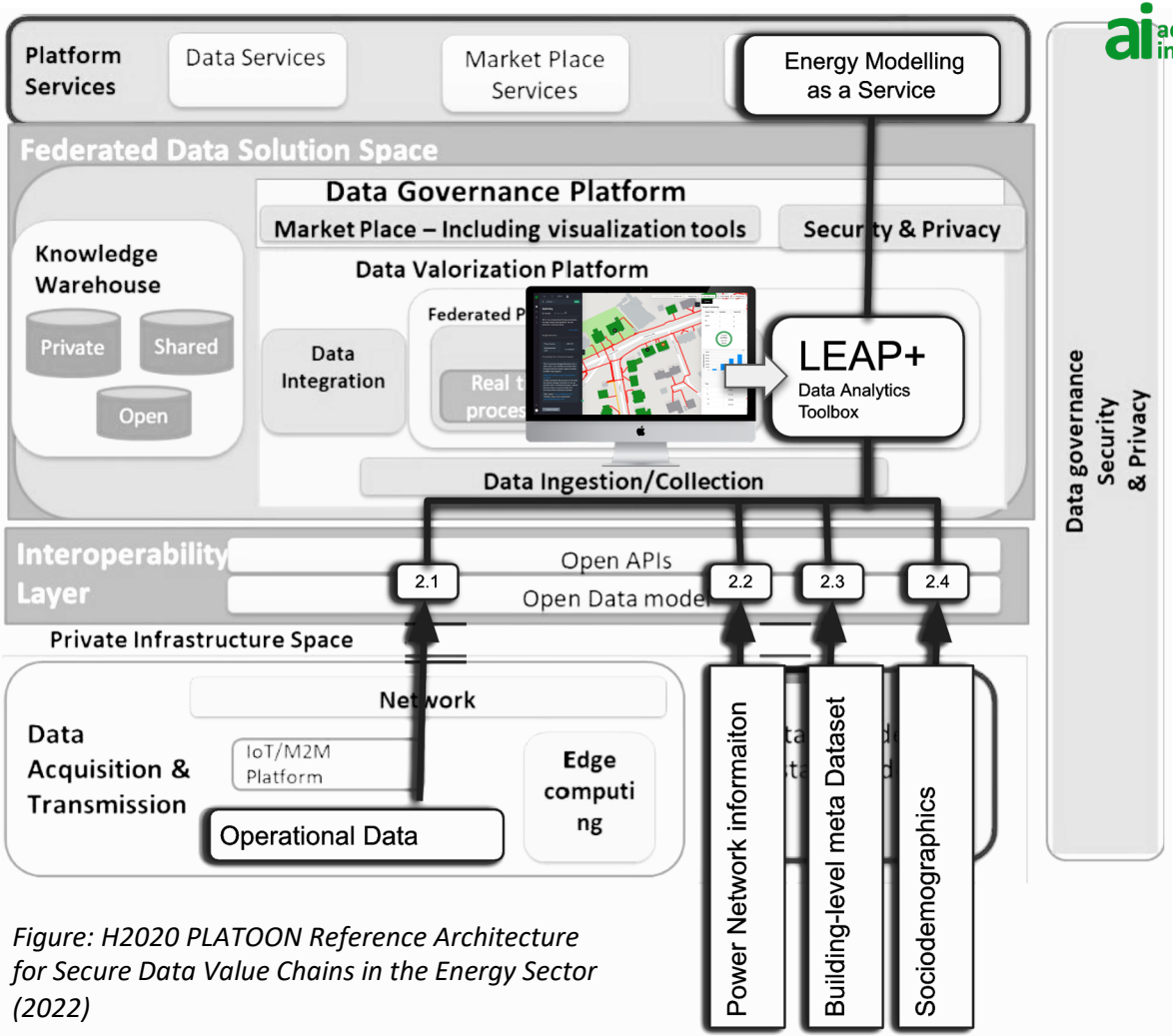


Figure: H2020 PLATOON Reference Architecture for Secure Data Value Chains in the Energy Sector (2022)

1. Technical scope
2. Selection of algorithms and tools
3. Scalability & flexibility
4. Data security and legal compliance
5. Quality assurance and Risk management
6. Degree of innovation of the technical solution
7. **Status of the mock-up & Product Demo**

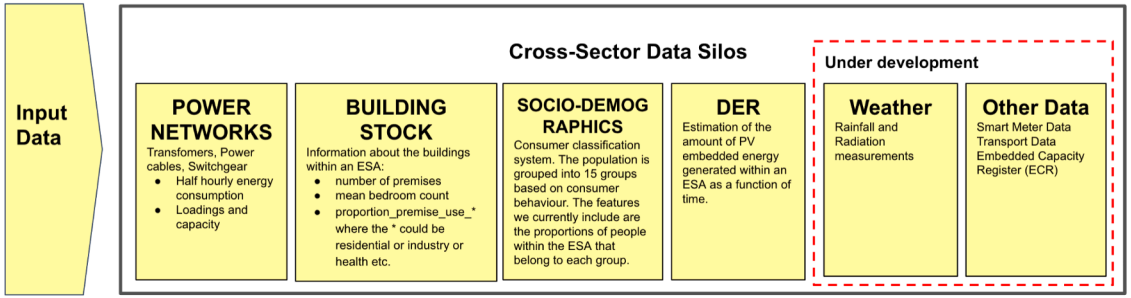
The demo will show:

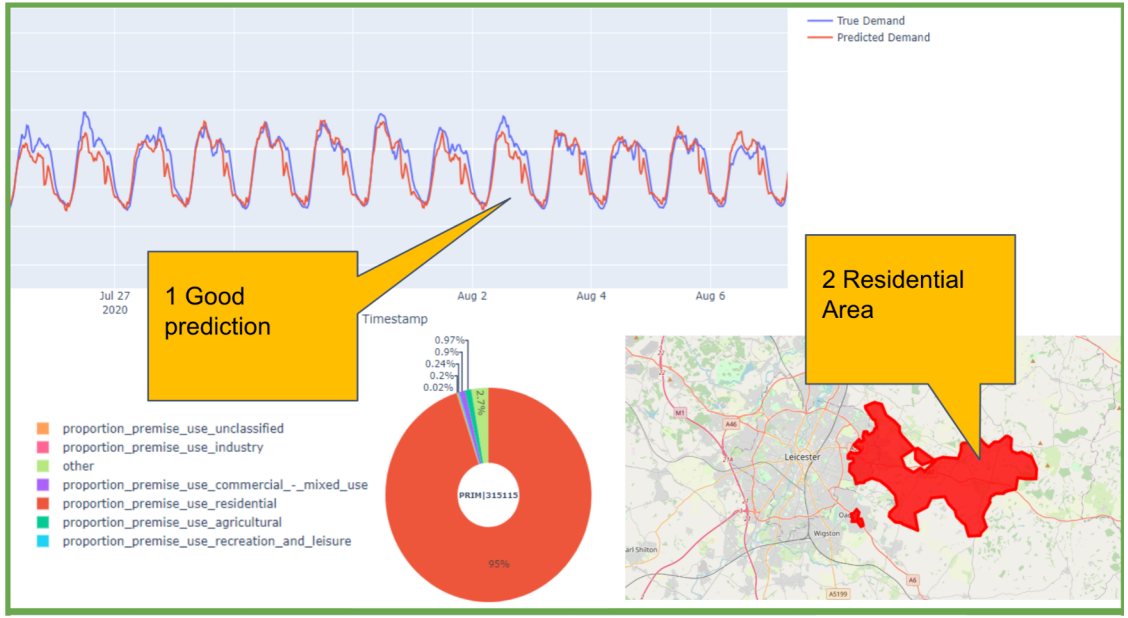
1. Visualisation platform
2. LCT Potential datasets
3. Use Case: Analysing a project and returning a cost and analysis

1. Technical scope
2. **Selection of algorithms and tools**
3. Scalability & flexibility
4. Data security and legal compliance
5. Quality assurance and Risk management
6. Degree of innovation of the technical solution
7. Status of the mock-up & Product Demo

The Data Analytics Component includes a model than can predict electricity demand at a given time for a range of different Electrical Supply Areas (ESAs) given some input features relating to the corresponding ESA.

The machine learning model is developed using GCP and xgBoost modls and details are in the annex. Current performance is described (bottom right)





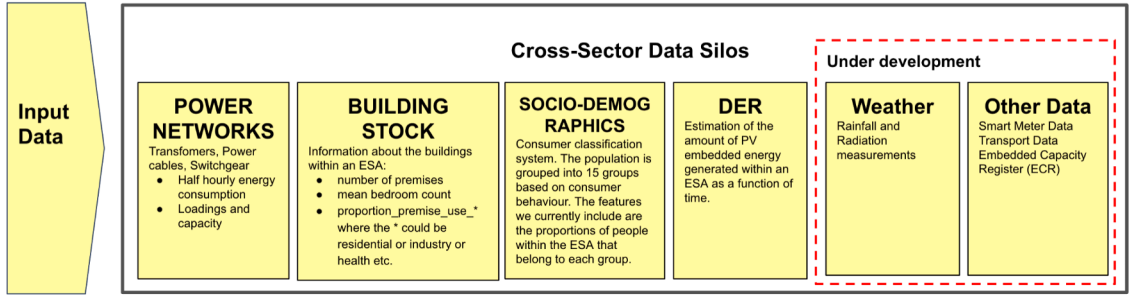
1. Technical scope
2. Selection of algorithms and tools
3. **Scalability & flexibility**
4. Data security and legal compliance
5. Quality assurance and Risk management
6. Degree of innovation of the technical solution
7. Status of the mock-up & Product Demo

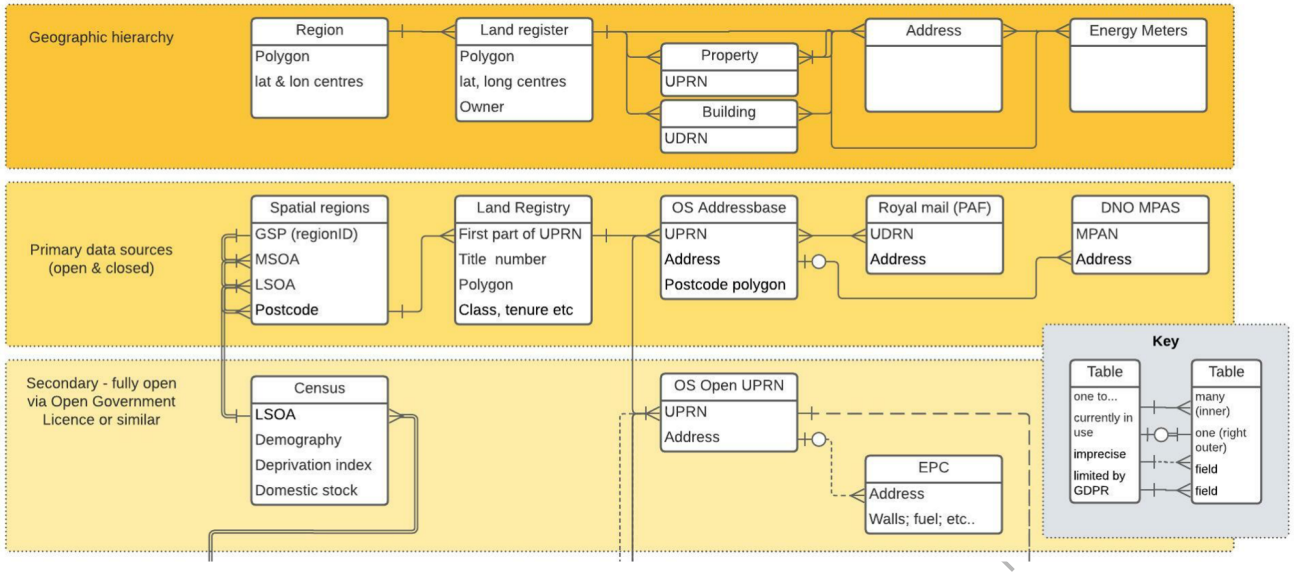
The Data Pipeline & Visualisaiton Components are comprised of an API gateway managing a series of lambda microservices that connect to data infrastructure of the service.

In certain cases, lambda functions are coupled to run sequentially as step functions (Data Ingest pipeline).

Since Lambda are either available on demand or as endpoints, they can be directly invoked through the user interface or sequentially as a part of a state orchestrated solution.

The solution has been designed to be interoperable with any standard third party data providers and service providers through API.





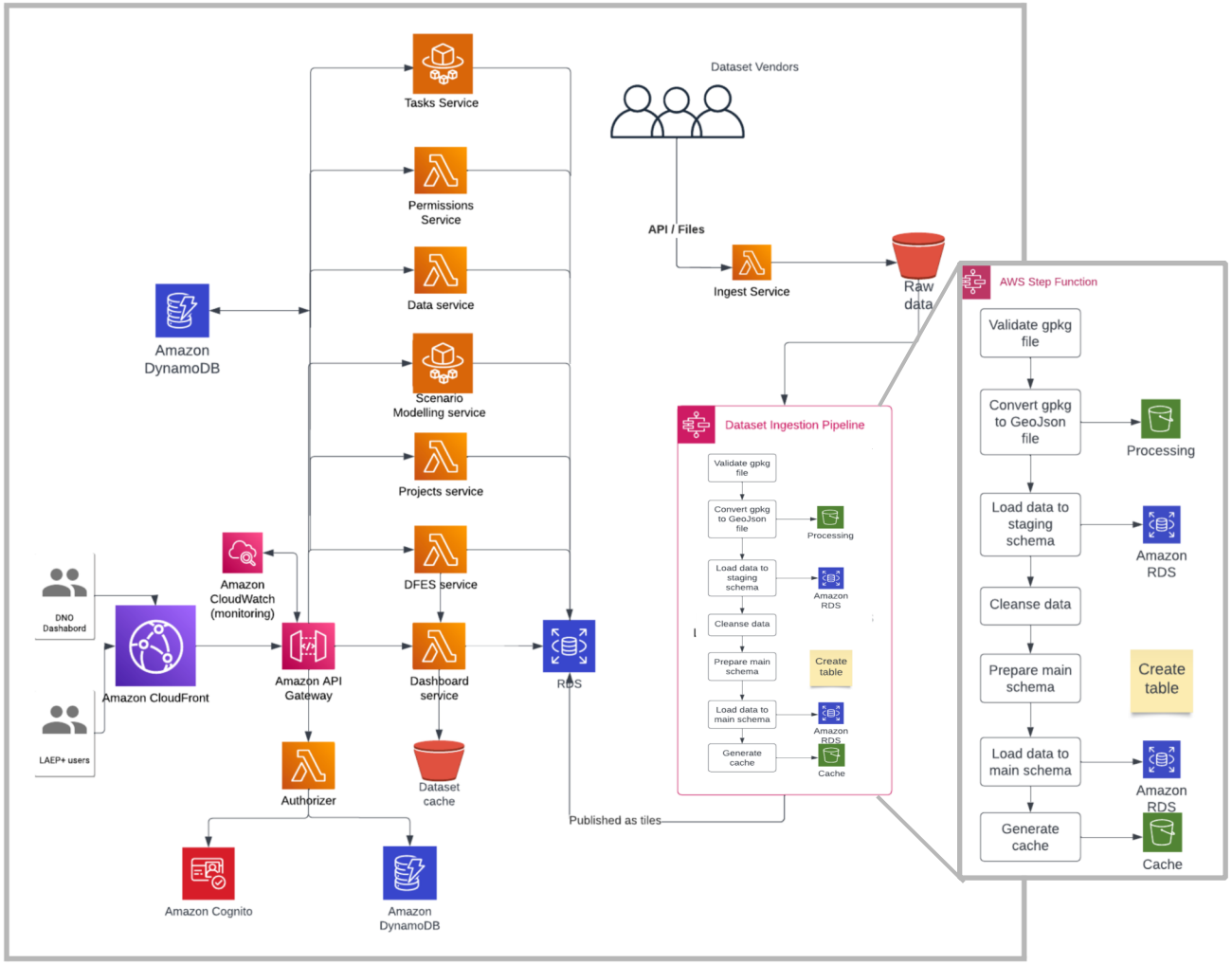
1. Technical scope
2. Selection of algorithms and tools
3. **Scalability & flexibility**
4. Data security and legal compliance
5. Quality assurance and Risk management
6. Degree of innovation of the technical solution
7. Status of the mock-up & Product Demo

The Data Pipeline & Visualisaiton Components are comprised of an API gateway managing a series of lambda microservices that connect to data infrastructure of the service.

In certain cases, lambda functions are coupled to run sequentially as step functions (Data Ingest pipeline).

Since Lambda are either available on demand or as endpoints, they can be directly invoked through the user interface or sequentially as a part of a state orchestrated solution.

The solution has been designed to be interoperable with any standard third party data providers and service providers through API.



1. Technical scope
2. Selection of algorithms and tools
3. Scalability & flexibility
4. **Data security and legal compliance**
5. Quality assurance and Risk management
6. Degree of innovation of the technical solution
7. Status of the mock-up & Product Demo

Smart meter energy consumption data at the building-level is classified as Personal Identifiable Information (PII). The Energy industry has a poor understanding of GDPR which prevents this data from being used effectively.

The loss of occupier and building information in the current anonymization process prevents the use of smart meter data by third party Energy Modelling Services. To address this challenge we will publish a template Data Anonymization Methodology for adoption by the industry.

Authentication and authorisation:
All users of the system will be required to use two factor authentication (MFA) to access data and will be required to provide KYC identification to allow access to sensitive data.

Compliance:
Some of our data sets will hold personally identifiable information that would be subject to GDPR regulation. We store strategically sensitive information that we have a duty of care to protect. If mishandled, data of this nature could be harmful.

Security:
All data on our systems is encrypted by default both in transit and at rest and complies with ISO 27000. The platform uses best practices for AWS account management using Organizations and Control Tower. Security guardrails in the form of Service Control Policies (SCP) ensure that AWS users (tech team) have the least privileged access. IAM policies and roles are leveraged to control access between AWS components - lambda, RDS and S3.

Ethics:
The responsible use of algorithms and data is paramount for the sustainable development of machine intelligence applications. We adhere to the UK's Digital Catapult Ethics Framework

Lawful Access:
We comply with IOC & GDPR lawful basis of access under legitimate interest. We will occupy the role of data processor.

1. Technical scope
2. Selection of algorithms and tools
3. Scalability & flexibility
4. Data security and legal compliance
5. **Quality assurance and Risk management**
6. Degree of innovation of the technical solution
7. Status of the mock-up & Product Demo

ISO 31000 Risk Management practices will be used throughout this project

A dynamic risk register utilising risk screening will be used to monitor the project, adhering to ISO 31000 principles

80% coverage for unit tests and development across three envinonemtns: DEV, SIT, & PROD

IT Security is in line with ISO 27000 and data is secured at rest and in transit

Business Continuity Plans are in place with back up locations and data recovery

Last Edit 25/05/2023			RISK REGISTER							
ID	TASKS	OWNER	OPPORTUNITY / RISK DESCRIPTION	PROBABILITY	IMPACT	RISK RATING	MITIGATION & CONTINGENCY	PROBABILITY	IMPACT	RISK RATING
ID	TASKS	OWNER	OPPORTUNITY / RISK DESCRIPTION	PROBABILITY	IMPACT	RISK RATING	MITIGATION & CONTINGENCY	PROBABILITY	IMPACT	RISK RATING
1	2	AITL	Agreement on UI & data requirements for Beta Release runs over, resulting in knock on delays to development.	3	4	7	AITL to supply detailed wireframes and specifications as part of Release Planning and Sprint Planning to accelerate discussion and common understanding. Target date for acceptance on requirements for core components & datasets is second week of the project.	2	2	4
2	2	AITL	ETL does not meet requirements to rapidly process and ingest datasets for all wards	5	4	9	Add the ETL requirements to sprint one. Requested one the Customer wide dataset from data team to test ETL to break down the given dataset to smaller datasets. Visualize at least 2 regions to demonstrate the gpkg breakdown and visualisation by end of Sprint 1.	2	4	6
3	2	AITL	Agreement of the feature specification and data flows for the Projects component takes time to finalise, resulting in a delay to component delivery	3	4	7	AITL to provide detail request for information in an RFI on project start and set up follow on meetings with key persons within the Customer to agree a specification and a timeline for component delivery	1	4	5
4	3	AITL	A customer success & support model between AITL<>the Customer takes time to agree, resulting in uncertainty that affects delivery of support and success components (Live Chat, User Guidance Documentation, Support Inbox, & User Activity Tracking) and on-boarding plan (webinar, training materials, drop in sessions)	3	3	6	AITL to propose a model early in project. Aim to have agreement on components by Spint 2. Aim to have agreement on onboarding plan by Spint 4	1	3	4
5	2	AITL	Time to agree the Customer branding exceeds expected timelines, resulting in delays to delivery of Beta Release	3	3	6	AITL to propose some wireframes for the Customer branding at project start. the Customer to share a brand guidelines pack on project start. Engage early in conversation about changes required.	1	3	4
6	4	AITL	Agreement on plan for user testing exceeds expected timelines resulting in a delay to user acceptance testing	3	3	6	AITL to propose a plan for user testing by end of spint 4 for the Customer's input and agreement. the Customer to provide clarification on any requirements they have for user testing, as well as any beta users planned for inclusion			0
7	2	AITL	Data ingest and update via API takes time to integrate resulting in delays to data integration	3	3	6	Prioritise API set up for a collection of Data Log datasets to de-risk early on and identify any issues with integration	2	2	4
8	2	AITL	DFES data handover is delayed, resulting in a delay to DFES integration	2	2	4	AITL to share Data Log and RFI at project start listing required datasets. the Customer to support with data collection and processing for the Customer owned datasets.	1	1	2
9	1	AITL	Communication difficulties due to dispersed team leading to misunderstandings	1	2	3	Effective use of IT collaboration tools and regular project meetings to ensure shared understanding. Use of Zoom, Confluence, Mural, GitHub and Jira collaboration and PM tools	1	2	3
10	1	AITL	Cost overruns exceed project budget and threaten delivery	1	2	3	Ensure that all costs are firmly quoted in early stage of project and reviewed at weekly Project Reviews. If overruns are identified inform senior management team and mitigate asap	1	1	2
11	2	AITL	Data visualisation of large dataset is not optimised for users (e.g. buildings are not help when looking across large areas, needs to aggregate) and can sometimes render poorly	4	3	7	Test data aggregation solution to display aggregated data at higher zoom levels	2	2	4

KEY OF TASK TYPES AND OWNERS

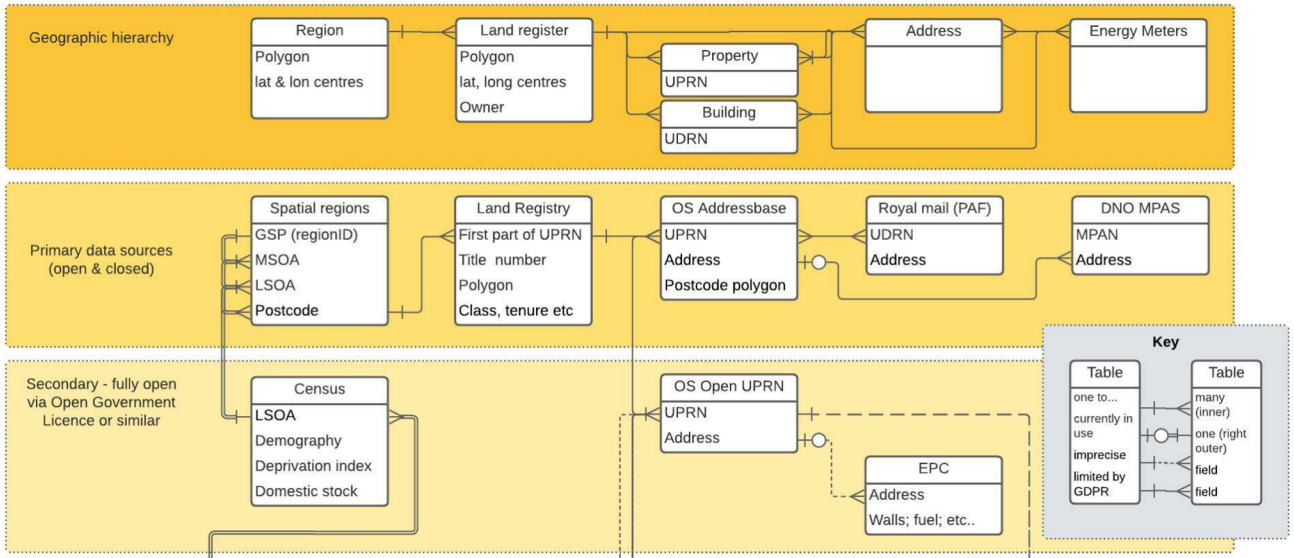
(Tech) Technical (PM) Project Management (Com) Commercial (UR) User Research (Env) Environment (Leg) Legal (HR) Human Resources (Reg) Regulatory

Figure (above): Dynamic risk register

1. Technical scope
2. Selection of algorithms and tools
3. Scalability & flexibility
4. Data security and legal compliance
5. Quality assurance and Risk management
6. **Degree of innovation of the technical solution**
7. Status of the mock-up & Product Demo

6.1 Data Management: The innovative solution utilises an automated data pipeline to manage and update datasets and models continuously at national scale (top). This is a significant challenge due to the lack of interoperability between data sources due to the absence of common identifiers, commercial barriers and licensing constraints.

6.2 Low-Carbon Technology (LCT) Potential is defined as the suitability of a building for retrofit (such as Building Fabric Upgrades, PV, EVCP, ASHP, GSHP.) The innovation calculates LCT Potential to PAS2035 standard. Every technology has a modelling process that is refreshed at the same rate as the source data. The example (bottom) is for heat pumps.



(above) 6.1 Data Management & Pipelines (below) 6.2 Low-Carbon Technology (LCT) Potential

ANNEXS

1. Technical scope
2. Selection of algorithms and tools
3. Scalability & flexibility
4. Data security and legal compliance
5. Quality assurance and Risk management
6. **Degree of innovation of the technical solution**
7. Status of the mock-up & Product Demo

6.1 Data Management: The innovative solution utilises an automated data pipeline to manage and update datasets and models continuously at national scale (top). This is a significant challenge due to the lack of interoperability between data sources due to the absence of common identifiers, commercial barriers and licensing constraints.

6.2 Low-Carbon Technology (LCT) Potential is defined as the suitability of a building for retrofit such as Building Fabric Upgrades, PV, EVCP, ASHP, GSHP. The innovation calculates LCT Potential to PAS2035 standard. Every technology has a modelling process that is refreshed at the same rate as the source data. The example (bottom) is for heat pumps.

22 FOR TRACK 3 ONLY: Please provide general information about the EXTERNAL dataset that you will use (i.e name, data provider, dataset categories and etc.) Teams are not permitted to solve the challenge only with open source data or their own startup's internal data - they must be partnering with an external data provider and use their external data. (??? char)

DATA PROVIDER 1 CleanTech
We will use a Distributed Energy Resources (DERs) map co-developed by Sheffield University. The geolocated point dataset is a register of over 1,000,000 DERs distributed throughout the UK representing 13GW of power generation capacity with half hourly output data.

DATA PROVIDER 2: CleanTech
We will use smart meter data provided by Grid Network Operator SSEN. The time series dataset (quarter hourly) energy consumption measured in kWh of up to 3 million electricity consumers associated with address data.

DATA PROVIDER 3: Cleantech
We will use the REACH dataset CERTH Energy information related to building energy consumption, generation and storage. The time series dataset (quarter hourly) energy consumption measured in kWh of up to 93 electricity consumers associated with environmental data.

DATA PROVIDER 4: PropTech
We will use building stock data provided by Ordnance Survey. The geolocated point dataset contains information on build with up to 500 attributes. Attributes include building characteristics (Energy Performance Rating, living space, number of bedrooms). This data will be linked to EPC & INSPIRE, an open source dataset of building energy performance & polygons issued under the European Green Deal Data Space.

DATA PROVIDER 5: FinTech
We will use socio demographics data provided by Experian MOSIAC commercial dataset, a leading EU credit referencing agency. The postcode-level dataset contains propensity scores for income and demographic including attributes such as car ownership, family composition and age. This data will be linked to open source census data.

D2.3 Address level estimation of HH demand

Model Schematic

Input Data

Cross-Sector Data Silos

POWER NETWORKS

Transfomers, Power cables, Switchgear

- Half hourly energy consumption
- Loadings and capacity

BUILDING STOCK

Information about the buildings within an ESA:

- number of premises
- mean bedroom count
- proportion_premise_use_* where the * could be residential or industry or health etc.

SOCIO-DEMOGRAPHICS

Consumer classification system. The population is grouped into 15 groups based on consumer behaviour. The features we currently include are the proportions of people within the ESA that belong to each group.

DER

Estimation of the amount of PV embedded energy generated within an ESA as a function of time.

Under development

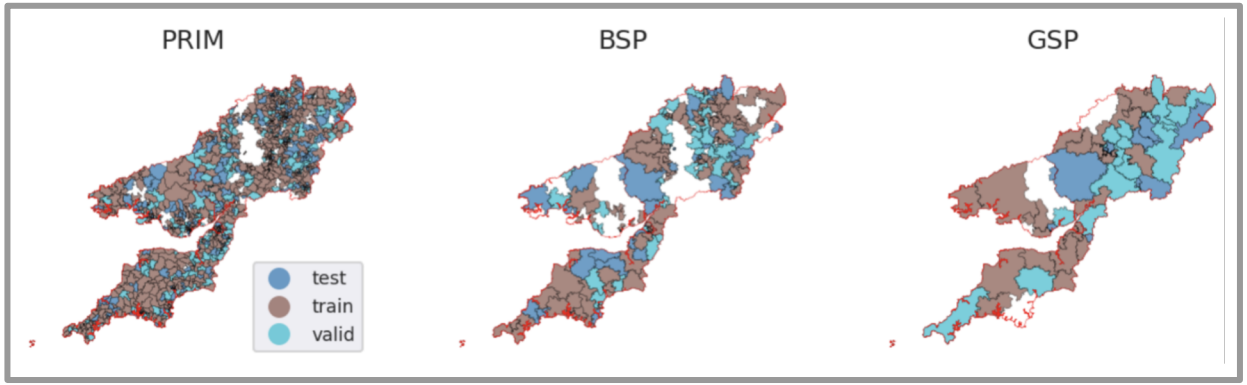
Weather

Rainfall and Radiation measurements

Other Data

Smart Meter Data
Transport Data
Embedded Capacity Register (ECR)

Labels (Training and Test Data)

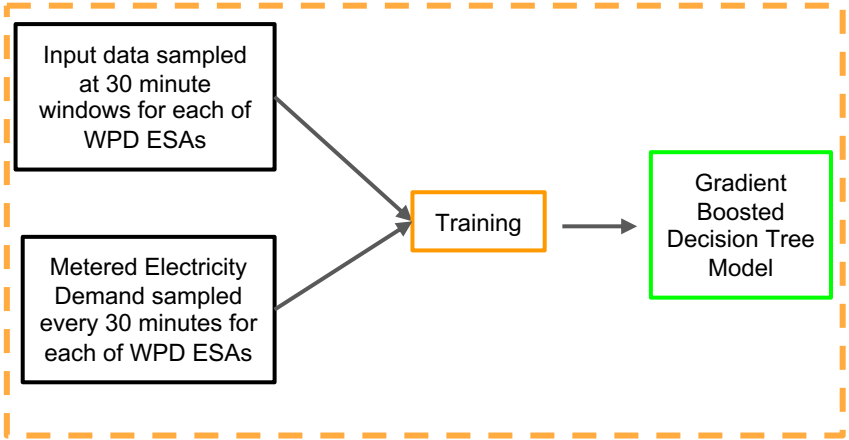


D2.3 Address level estimation of HH demand

Model Training

Step 1: Training

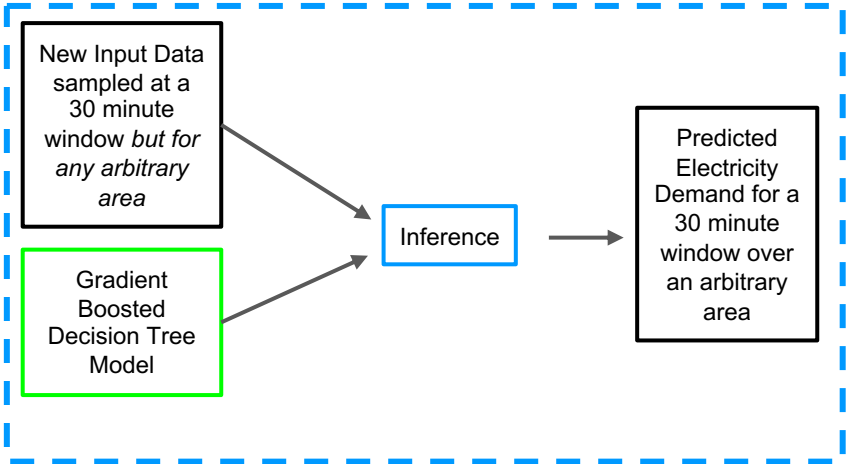
We have to use ESAs to train the model because they are the only areas we have “true” values for the electricity demand.



Step 2: Inference

The model is capable of predicting the electricity demand for any area that input data can be collated over. **The area does not need to be an ESA.**

Computationally cheap and therefore the model can easily predict weeks, months or years of demand for an area.



D2.3 Address level estimation of HH demand

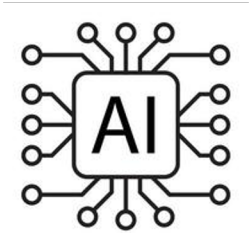
Future commercial deployment of the model

1. An area is identified (maybe by drawing a polygon or choosing a postcode). A timeline is also chosen: I.e. 01/01/2020 - 01/06/2020

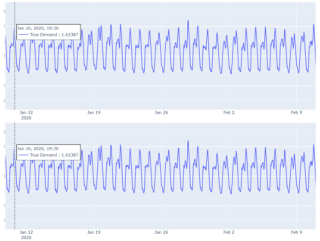
2. The input data for that area and within the desired timeline is extracted from a database. This will probably require interpolating some of the data sources to the desired area.

Database

3. Feed data into the model and predict the electricity demand profile for the requested timeline.



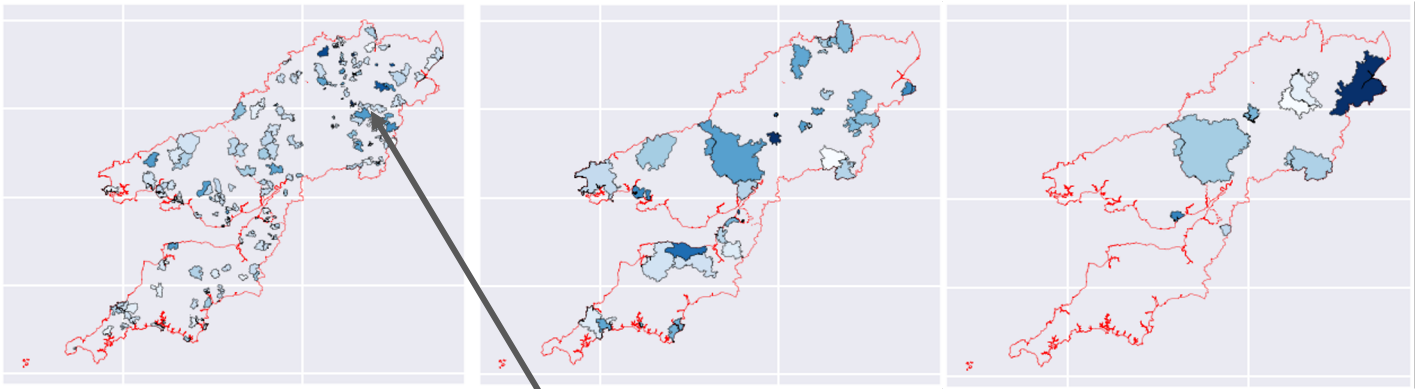
4.Predicted Historical Electricity Demand Profile

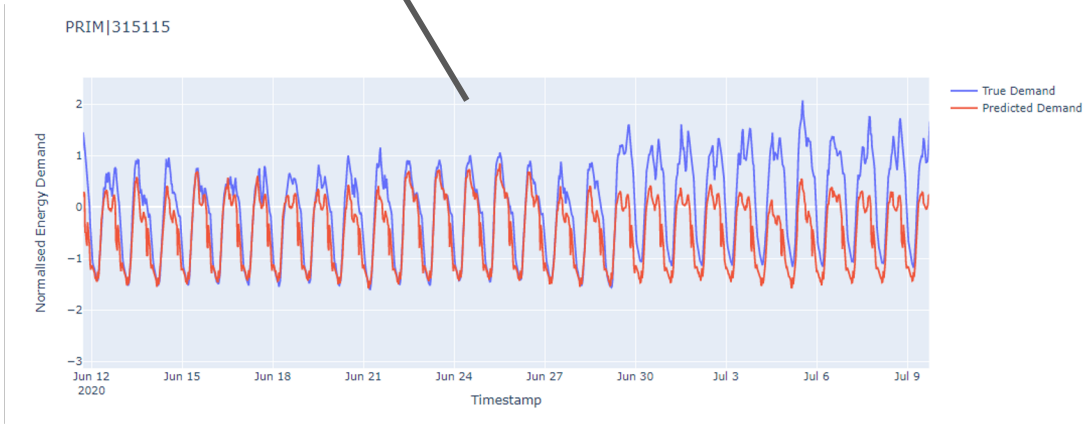


D3.1 Validation complete against substation data

Model performance at accross increasing spatial scales

Primary BSP GSP





Model Performance across different scales

On the maps, the darker blue, the higher the error. Both these plots show results on “test” ESAs. These are ESAs that were not used to train the model.

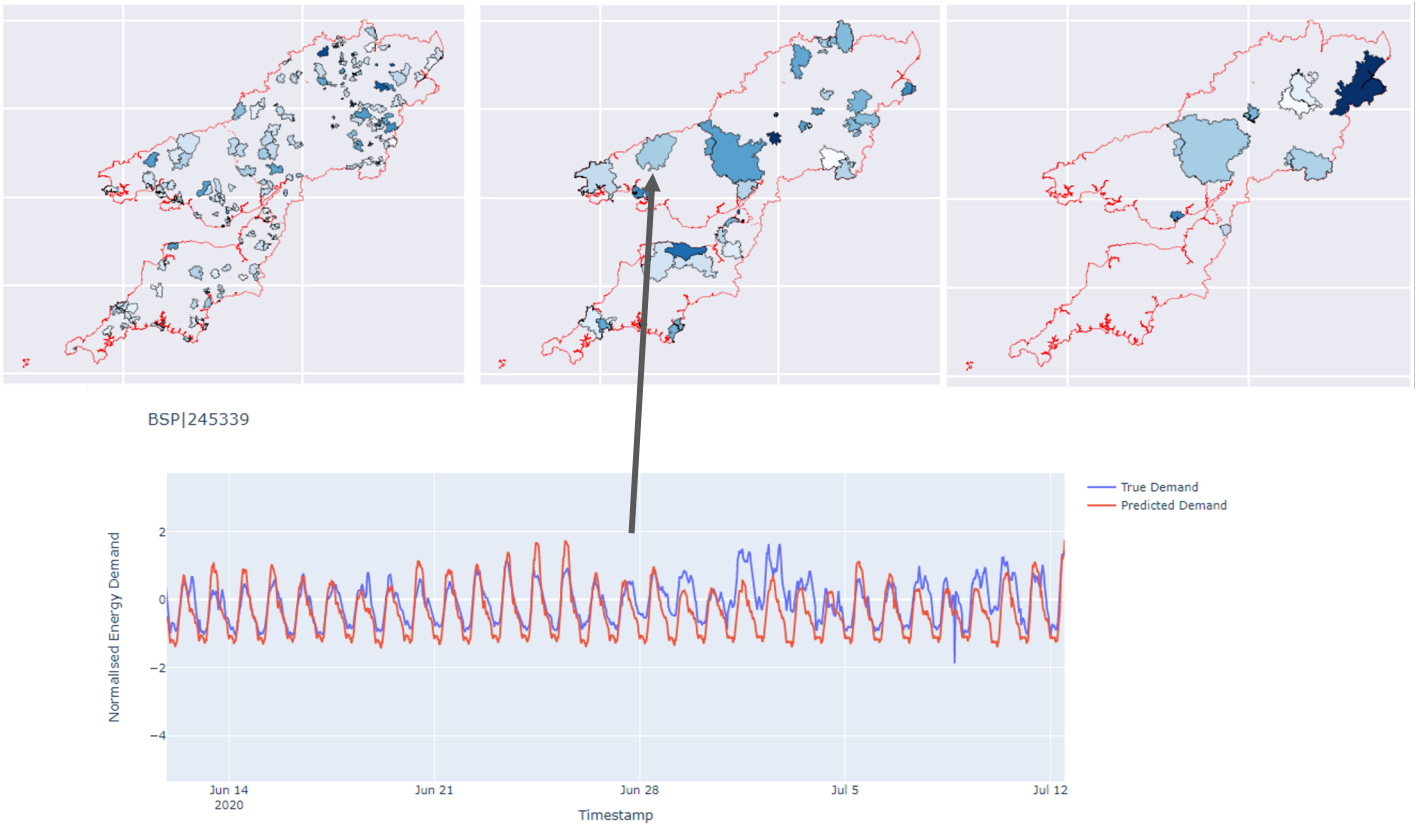
D3.1 Validation complete against substation data

Model performance at accross increasing spatial scales

Primary

BSP

GSP



Model Performance across different scales

On the maps, the darker blue, the higher the error. Both these plots show results on “test” ESAs. These are ESAs that were not used to train the model.

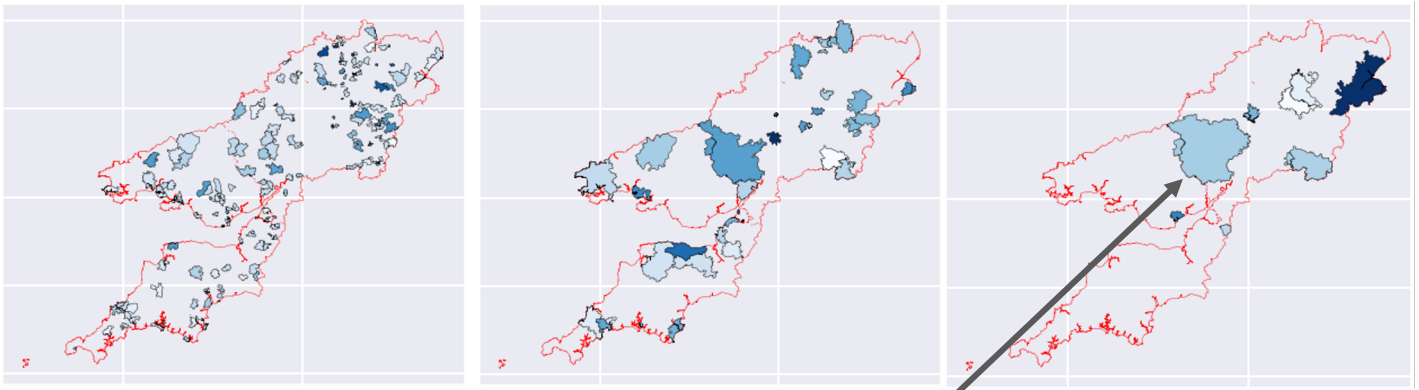
D3.1 Validation complete against substation data

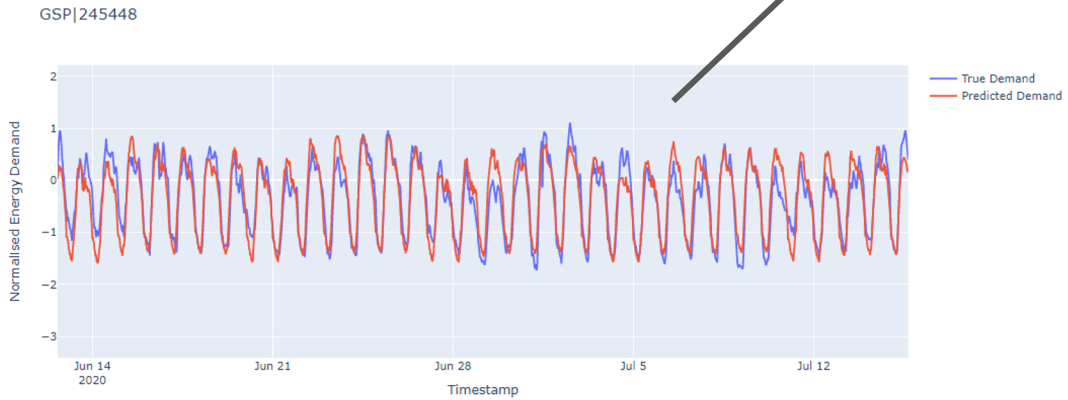
Model performance at accross increasing spatial scales

Primary

BSP

GSP





Model Performance across different scales

On the maps, the darker blue, the higher the error. Both these plots show results on “test” ESAs. These are ESAs that were not used to train the model.

D3.1 Validation complete against substation data

Model performance at across demographic ranges

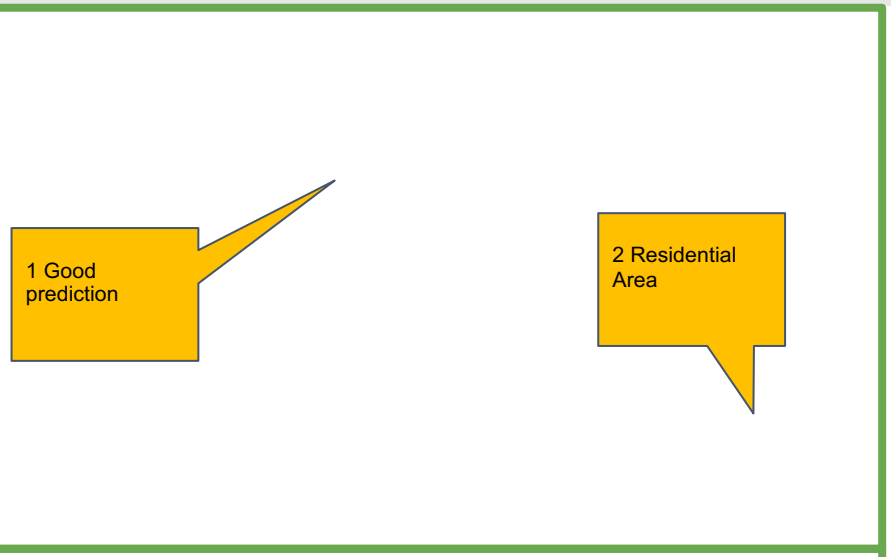
Residential vs Industrial Regions:
Areas that are dominated by residential usage are typically better predicted by the model. Areas that have larger proportions of industrial usage are more challenging to model.

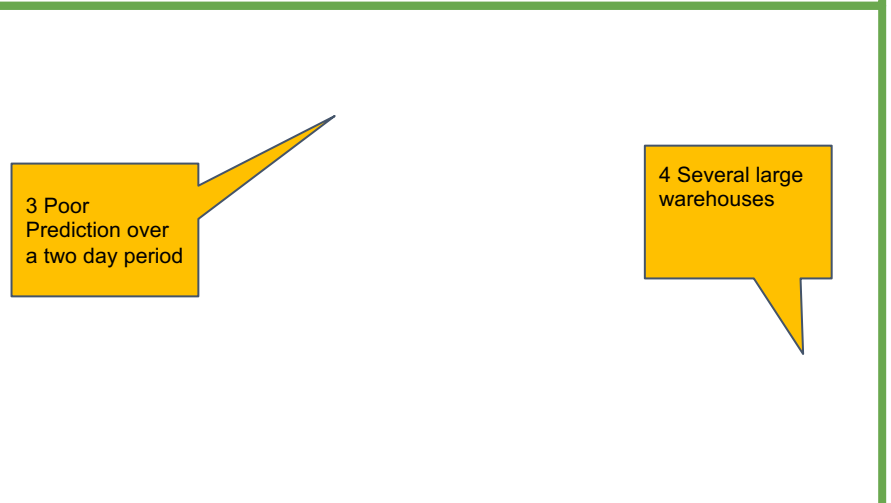
Most Important Inputs:

Energy generated from solar

Date/Time features are highly important

Group_c and group_n are proportions of people in rural and urban areas respectively

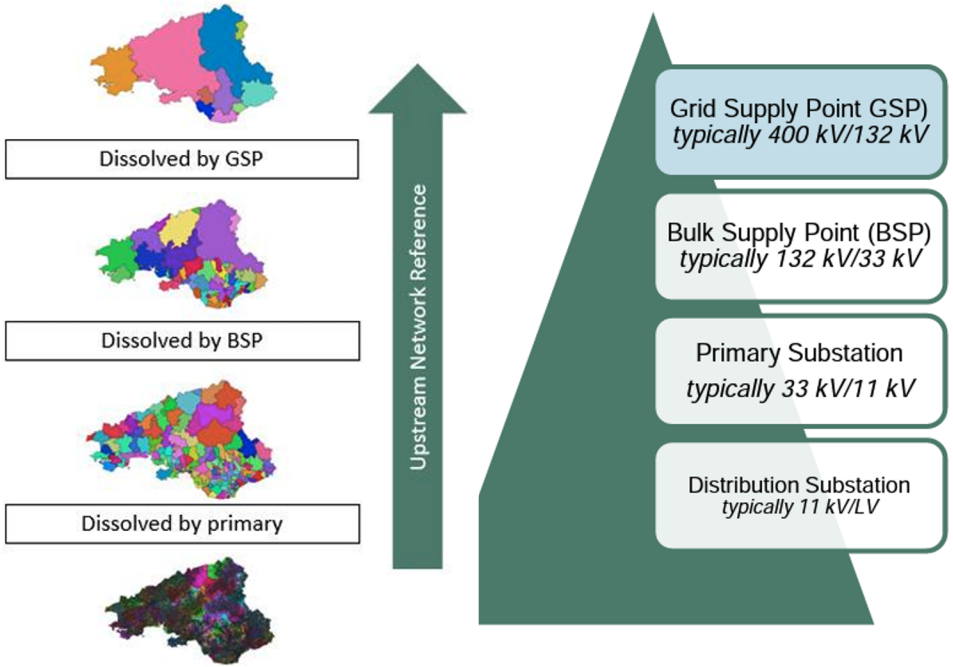




Annex (recall from QRM1)

As part of the transition towards being a Distribution System Operator (DSO), it is necessary to be able to translate between datasets which describe the distribution network assets and geographic datasets for various modelling purposes. DNO use Electricity Supply Areas in the following DSO activities:

1. **Distribution Future Energy Scenarios** – used to spatially allocate demand and generation growth to the most likely point of connection to the distribution network. The level of ESAs used in DFES studies are determined by the level of electrical analysis that is undertaken.
2. **Network capacity map** – used as a layer on the network capacity map to demonstrate the geographic area supplied by a Primary substation, Bulk Supply Point or Grid Supply Point.
3. **Network flexibility map** – used to visualise the Constraint Management Zones where DNOs procures flexibility services. A postcode to Primary substation lookup is also used as part of the network flexibility map



Generating Voronoi polygons around transformer centroids

To produce the Electricity Supply Areas, the input datasets are combined and substations (HV/LV) within a licence area . An algorithm is run to create filtered to all distribution Voronoi or each . Each polygon represents an area which has a common distribution substation as the closest point of connection to the HV network. distribution substation Figure 3 polygons f visually demonstrates how each distribution substation has a small Voronoi polygon.

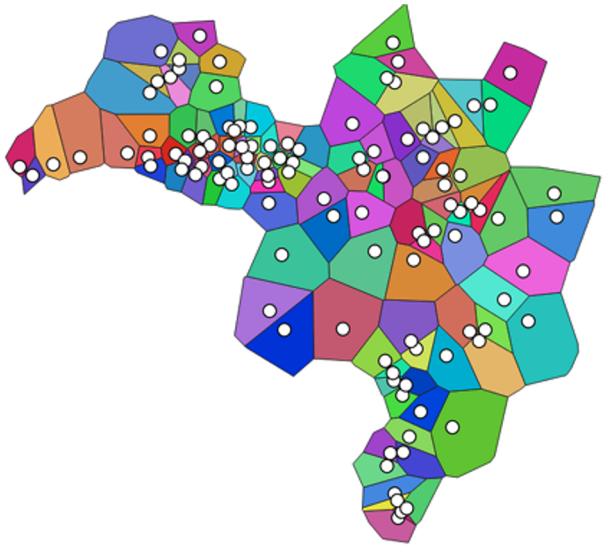
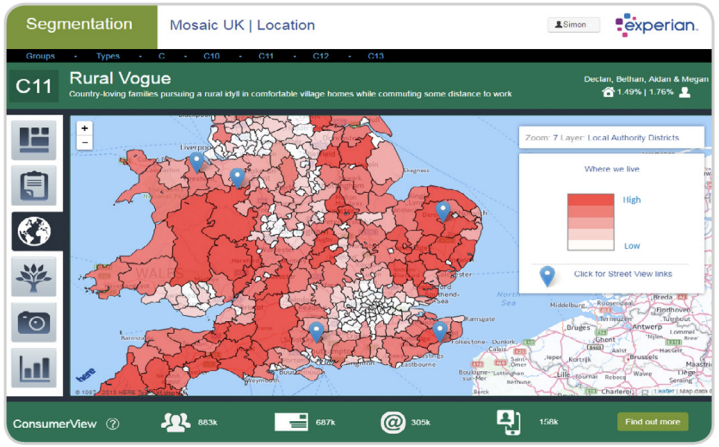


Figure 3 - Bingham primary (East Midlands) Voronoi polygons

Annex (recall from QRM1)

Mosaic is Experian's most comprehensive cross-channel consumer classification, built for today's hyper-connected world. Using new data and analytical methods, you get deeper insights on consumer lifestyles and behaviour to help you make more informed marketing decisions

Group	Description	Type	Description
A	High status city dwellers living in central locations and pursuing careers with high rewards	A01	Global high flyers and moneyed families living luxurious lifestyles in the most exclusive boroughs (No type A01 in Nottinghamshire)
		A02	High status households owning elegant homes in accessible inner suburbs where they enjoy city life in comfort
		A03	City professionals renting premium-priced flats in prestige central locations (No type A03 in Nottinghamshire)
		A04	Career-minded 20 and 30-somethings renting expensive apartments in highly commutable areas of major cities (No type A04 in Nottinghamshire)
B	Established families in large detached homes living upmarket lifestyles	B05	Asset-rich families with substantial income, established in distinctive, expansive homes in wealthy enclaves
		B06	Retired residents in sizeable homes whose finances are secured by significant assets and generous pensions
		B07	High-achieving families living fast-track lives, advancing careers, finances and their school-age kids' development
		B08	Well-off families in upmarket suburban homes where grown-up children benefit from continued financial support
		B09	Mature couples in comfortable detached houses who have the means to enjoy their empty-nest status



Annex (recall from QRM1)

UKBuildings is a unique spatial property database of building characteristics created by Verisk. The dataset offers information about individual buildings such as the age, use, height, residential type, number of floors, location, structural detail, or area of a building. The core UKBuildings product has been created from the observation of aerial imagery and published open data ensuring that all buildings, both residential and non-residential, even those without addresses, are classified. This dataset is licenced for RESOP and LEO projects under a 6 month pilot.

Identify Results

Feature	Value
premise_age	Early/middle Victorian1837 - 1870
premise_use	Residential
premise_type	Large semi detached
premise_floor_count	2
height	10.9
premise_area	93.2
building_area	186.5
address_area	186.5
gross_area	186.5
basement	0
listed_grade	
element_count	1