

Energy Data Science Collaboration Forum

In partnership with the ENA

Terms of Reference for:

<https://github.com/Energy-Data-Science-Collaboration-Forum>

Version	Author	Position	Approver	Position	Date
1.0	Gabriel Griffin-Booth	Data Science Manager, Gas Transmission and Metering	Matt Webb	ENA DDSG Chair; Head of Enterprise Data Management, UKPN	21/07/2022

Introduction

Companies in the GB Energy Industry are hiring data scientists. These data scientists are asked to model similar phenomena seen on the GB Energy System; however, their companies may not be direct competitors. Academia also improves models of the same phenomena, albeit with lesser access to industry data. Separate activities within the ENA Data and Digitalisation Steering Group (DDSG) are exploring wider access to data and is out of the scope of this proposed forum, although it is intended to align with the broader open data agenda that is developing within the sector.

The common modelling activities from these companies present an opportunity for data science collaboration and idea sharing, such that models used in business decision making can be improved. This coordinated collaboration across the energy sector can be to the benefit of both customers and stakeholders, for example enhancing tools that promote more efficient operation and management of GB energy system has the potential to reduce pass through costs to the end consumer.

Examples of models that companies make in parallel are demand forecasts for electricity and gas, and weather-based wind and solar generation forecasts. The uncertain future electrification of domestic heating and the addition of grid scale battery balancing assets as an alternative to natural gas fired plant to balance unreliability in renewable output highlights a future interaction between electricity and gas demand models that has the potential to be explored. These models are continuously under development and improvement in academia and have been the subject of recent NIA projects to aid decision making, and further collaboration in this area presents an opportunity to help in the co-development of future SIF and NIA projects.

Companies make asset investment decisions; asset degradation predictions based on usage levels could be improved by the sharing of best practice in forecasts of energy demand across GB, especially where the companies in question are not direct competitors e.g., Electricity line vs Gas pipe asset owners.

Purpose and Aim

The purpose of this forum is to create a framework for Data Science teams to collaborate in a transparent and controlled manner with other companies in the Energy Sector, whilst being able to signpost the structure to their respective leadership when seeking authorisation to publish company code as open source. The aim is to enable the sharing of best practice or techniques via creating a meeting and open code sharing space to enable Data Scientists to "talk shop". This may be particularly attractive for companies who are not direct competitors but require the creation of similar models to more efficiently operate or manage aspects the GB energy system, and therefore be willing to open source publish elements of models being used.

This forum also provides a route for Academia to understand the modelling approaches used in the industry, with the aim of enabling academics to contribute improvements to techniques and reducing an observed lag at the start of innovation projects where academics must first understand how the industry is working and modelling, which enables a stronger outcome of the innovation projects. It is also intended that these knowledge sharing discussions lead to the generation of NIA/SIF project proposals.

Energy Data Science Collaboration Forum

In partnership with the ENA

Inputs

The forum would be open to input from Data Scientists across electricity & gas transmission and distribution asset owners and system operators, and associated organisations such as the Energy Catapult.

The forum would be open to input from Academics with an interest in collaborating with data science teams, especially those with an interest in follow on NIA/SIF projects who wish to understand existing modelling approaches in industry or who may want to propose improvements to existing models.

The input would take the form of sharing best practice and open-source publishing code.

Outputs and Success Criteria

The output will be an open-source code repository, containing code published under an open licence e.g. MIT. This could be the cross sharing of models, packages, or even just code snippets between teams following discussion. Some access control and quality rules (e.g. clear licencing terms included) will need to be managed to keep the repository in good order.

The success criterion for this would be the existence of a code repository (<https://github.com/Energy-Data-Science-Collaboration-Forum>) containing open source published code from more than one contributor, and the addition to this repository spurred by the ad hoc meeting of Data Scientists and Academics. Performance measures of the number and frequency of posts by discrete users and evaluating interactions with posts present opportunities to measure cycles of engagement and identify long term trends.

A secondary success criterion could be NIA/SIF project ideas that are spun out of this collaboration, or where the available code base enables the more efficient running of the project due to pre-established modelling knowledge shared.

Success metrics could be created by identifying a performance metric of a particular model and demonstrating its comparative improvement. e.g., improvement in demand forecasting following model changes, or the comparative performance metrics of different forecasting model approaches.

Frequency

The primary purpose is to enable ad hoc meetings and discussion between teams as per their availability and workload, however it would be prudent to schedule a check-in meeting once a quarter to monitor progress and highlight issues.