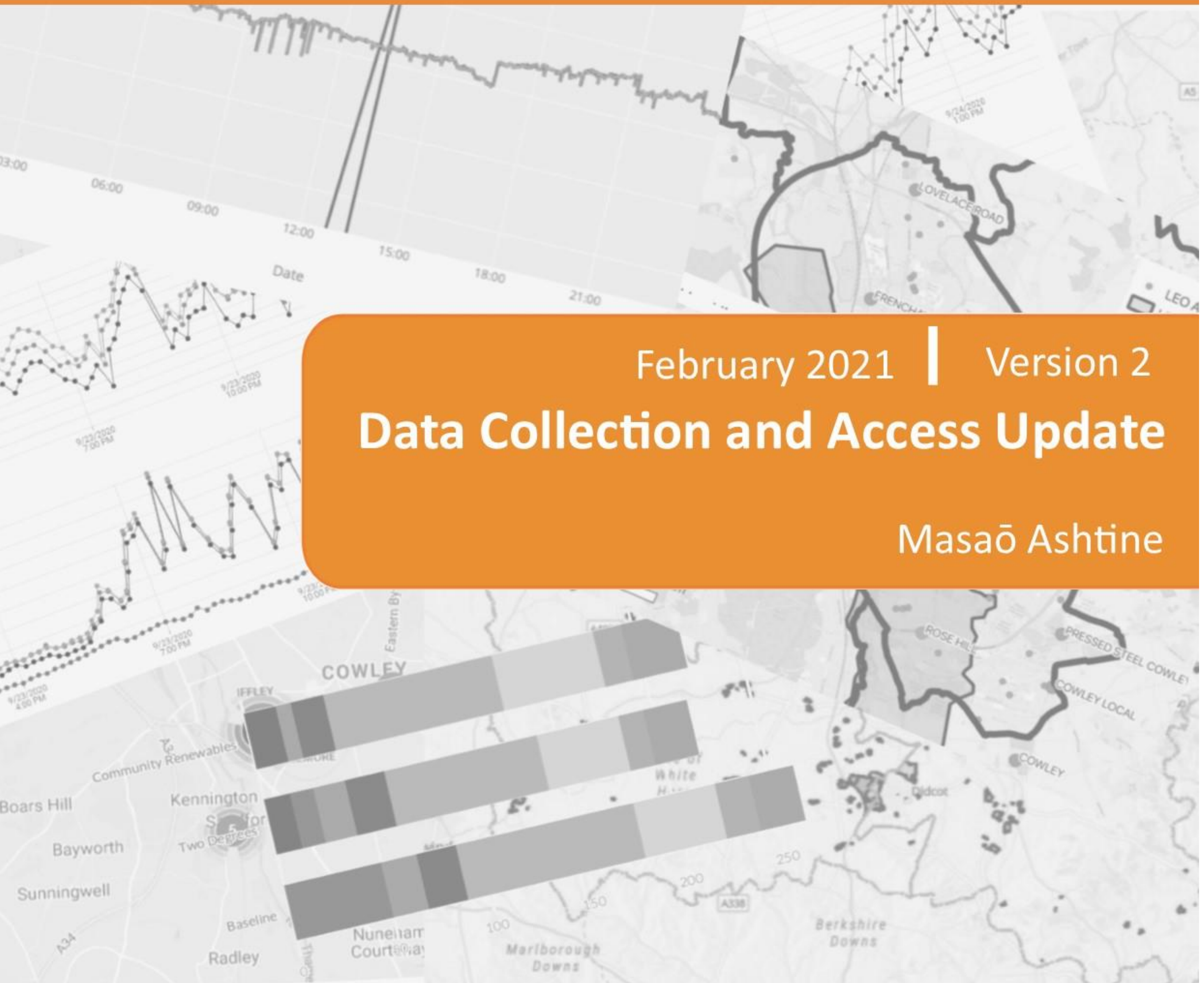
LEO Local Energy Oxfordshire



February 2021 | Version 2

Data Collection and Access Update

Masaō Ashtine

University of Oxford · Oxford Brookes University · low carbon hub · Oxford City Council · Oxfordshire County Council · Scottish & Southern Electricity Networks · piclo · EDF Energy · Nuvve · Industrial Strategy

Report Title:	Data Collection and Access Update Report
Author(s):	Masaō Ashtine
Organisation(s):	University of Oxford

Version:	2.0	Date:	28/01/2021
Workpack*:	WP4	Deliverable:	4.14
Reviewed by:	Rajat Gupta		
Date:	08/02/2021		
Signed off by:	David Wallom		
Date:	09/02/2021		

Can be shared (Y/N):	Internally	Y	Publicly	Y

Context

The UK Government has legislated to reduce its carbon emissions to net zero by 2050. Meeting this target will require significant decarbonisation and an increased demand upon the electricity network. Traditionally an increase in demand on the network would require network reinforcement. However, technology and the ability to balance demand on the system at different periods provides opportunities for new markets to be created, and new demand to be accommodated through a smarter, secure and more flexible network.

The future energy market offers the opportunity to create a decentralised energy system, supporting local renewable energy sources, and new markets that everyone can benefit from through providing flexibility services. To accommodate this change, Distribution Network Operators (DNOs) are changing to become Distribution System Operators (DSOs).

Project Local Energy Oxfordshire (LEO) is an important step in understanding how new markets can work and improving customer engagement. Project LEO is part funded via the Industrial Strategy Challenge Fund (ISCF) who set up a fund in 2018 of £102.5m for UK industry and research to develop systems that can support the global move to renewable energy called: Prospering From the Energy Revolution (PFER).

Project LEO is one of the most ambitious, wide-ranging, innovative, and holistic smart grid trials ever conducted in the UK. LEO will improve our understanding of how opportunities can be maximised and unlocked from the transition to a smarter, flexible electricity system and how households, businesses and communities can realise the benefits. The increase in small-scale renewables and low-carbon technologies is creating opportunities for consumers to generate and sell electricity, store electricity using batteries, and even for electric vehicles (EVs) to alleviate demand on the electricity system. To ensure the benefits of this are realised, Distribution Network Operators (DNO) like Scottish and Southern Electricity Networks (SSEN) are becoming Distribution System Operators (DSO).

Project LEO seeks to create the conditions that replicate the electricity system of the future to better understand these relationships and grow an evidence base that can inform how we manage the transition to a smarter electricity system. It will inform how DSOs function in the future, show how markets can be unlocked and supported, create new investment models for community engagement, and support the development of a skilled community positioned to thrive and benefit from a smarter, responsive and flexible electricity network.

Project LEO brings together an exceptional group of stakeholders as Partners to deliver a common goal of creating a sustainable local energy system. This partnership represents the entire energy value chain in a compact and focused consortium and is further enhanced through global leading energy systems research brought by the University of Oxford and Oxford Brookes University consolidating multiple data sources and analysis tools to deliver a model for future local energy system mapping across all energy vectors.

Table of Contents

1. Executive Summary[1]

This report outlines the main processes and standards involved in the collection and management of shared data (and associated metadata) within Project LEO (LEO). The content is an update of the 2019 report on the LEO Data Sharing Log where information has been added on current work being done to improve data access and processing. These updates revolve around two markedly different changes to the previous report. Firstly, initial findings that were presented focused only on data collection where this current report extends to data access and analysis. Secondly, the current version of this report places much more attention on the tools needed to explore data within LEO, both internally through project partners, and externally amongst various stakeholders. The report will first explore the previous system and what recent changes have been implemented to enhance the overall data experience within LEO. Many of the tools and databases used within LEO are also discussed, including near-term plans to migrate LEO's data to an appropriate Microsoft cloud database such as Azure, capitalising on the resources within Power BI to allow LEO partners greater access and analytical tools within our project. Finally, this report briefly touches upon how these changes can be adopted by *fast-followers* in other local energy systems to facilitate effective data management.

2. First Implementation

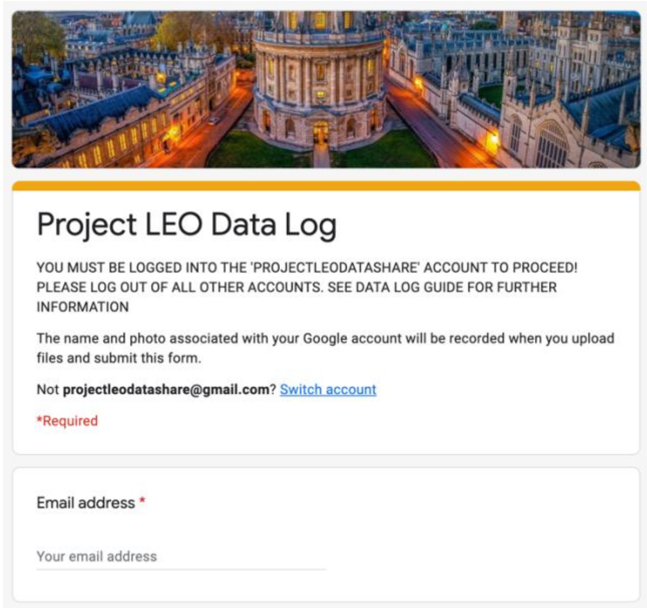
LEO's data can be categorised into two main streams: Foreground data (data produced within LEO's activities) and Background data (data sourced from databases external to LEO's direct work). Foreground data are of main focus within LEO's data management processes whereby data are stored in the project after being shared by partners. Background data however are processed for associated metadata and catalogued within LEO to allow partners the opportunity to explore datasets external to LEO that have been shared with partners, or, to access data to support their work within LEO.

2.1. Data collection

Foreground data are largely comprised of datasets associated with MVS (Minimum Viable System) trials have been conducted by project partners, and these datasets take on a life of their own in terms of reporting and data collection. Thus, LEO captures data differently from Foreground versus Background data, but both are securely logged and uploaded through the same online form.

[1] *Note, this report does not outline the detailed steps used to handle data in LEO, but the main processes and tools implemented within LEO, including links to other useful documentation. All proposed data platforms will be separate from the Integrated Land Use Mapping tool that has also been developed in LEO.*

When a partner logs data that have been shared within LEO (or externally) the Data Sharing Log is used.


Project LEO Data Log

YOU MUST BE LOGGED INTO THE 'PROJECTLEODATASHARE' ACCOUNT TO PROCEED! PLEASE LOG OUT OF ALL OTHER ACCOUNTS. SEE DATA LOG GUIDE FOR FURTHER INFORMATION

The name and photo associated with your Google account will be recorded when you upload files and submit this form.

Not **projectleodatashare@gmail.com**? Switch account

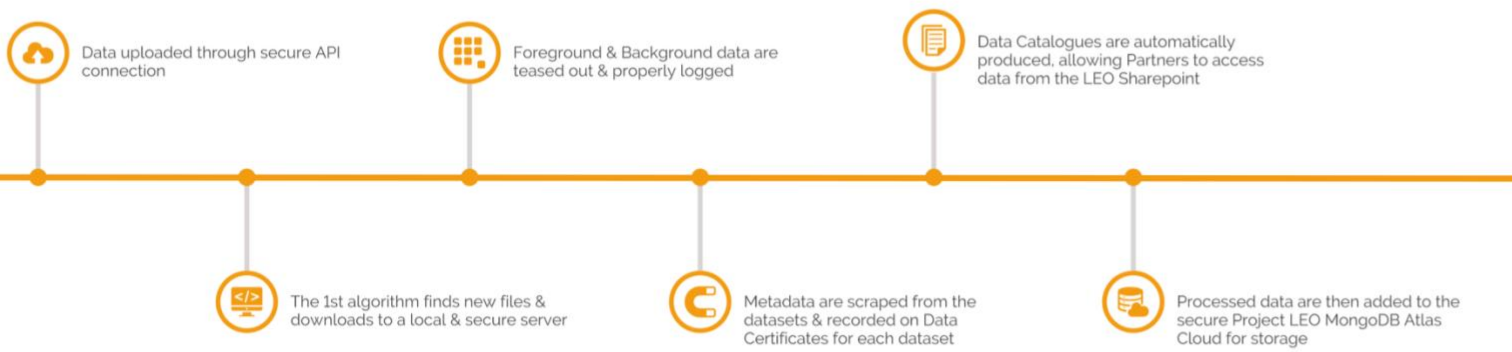
*Required

Email address *

Your email address

Google's services are used primarily within the data capture stages whereby a Google Form allows partners to meticulously log data and associated metadata, and Google's Drive functionality is utilised to store the data in this process. A bespoke form can be developed in LEO to handle the first point of contact with data, but we found this to be both resource and cost ineffective given the secure manner in which Google's tools can be easily accessed and utilised. Relying on widely adopted tools such as those found in Google's suite of software, also enhances LEO's appeal to external stakeholders and *fast-followers* (a term used in LEO to describe stakeholders that are able to quickly adopt methods and learnings within LEO to accelerate other local energy systems).

The ease at which data can be logged from mobile and desktop devices, as well as with datasets of large sizes, LEO's online form allows for increased flexibility and resources to our partners. From left to right, the timeline below shows how Project LEO's data are handled when partners use the Data Sharing Log to submit data.

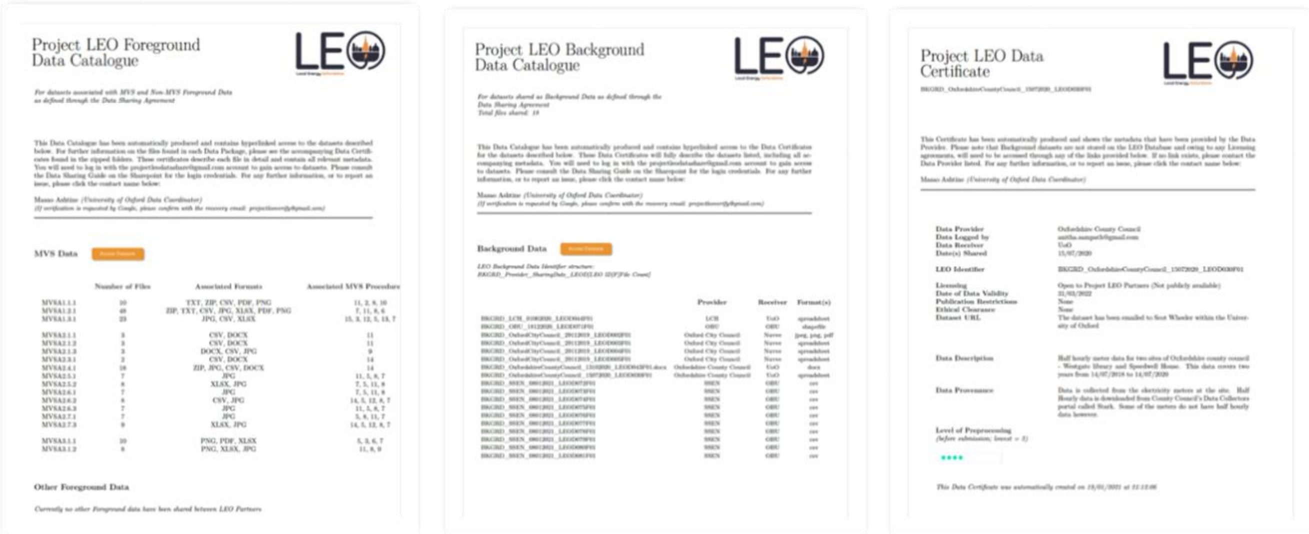


From left to right, LEO's data are first captured through an online form which, through and API (Application Programming Interface), allows data managers to move data to the LEO database after automatically scraping and processing metadata for both Foreground and Background datasets.

2.2. Backend processing

Once data have been logged by partners and uploaded to the LEO Drive account, these datasets are scraped for their metadata (producing Data Certificates and Catalogues as shown on the following page) and transferred to the LEO MongoDB cloud database for long-term storage. The use of MongoDB will be discussed in Section 3.2 as this service no longer stands as a cost-effective solution.

There are various scripts (Python) and APIs (Application Programming Interfaces) that are used to pull data from the LEO Drive, process the datasets and scrape their metadata, produce proper documentation in line with the FAIR principles, and transfer the data to the MongoDB database. These scripts and their accompanying documentation can be found on the LEO Repository (email us here if you are external to Project LEO and would like to learn more).


Project LEO Foreground Data Catalogue

MVS Data

Other Foreground Data

Project LEO Background Data Catalogue

Background Data

Project LEO Data Certificate

Only one Foreground and Background data catalogue (each) exists within LEO's database whereas each dataset and data file are accompanied by a Data Certificate.

2.3. Access

Presently (at time of reporting), the LEO database on MongoDB does not allow for easy access of data by partners. Foreground data catalogues give partners the opportunity to access data from LEO's Drive database, but this presents a number of issues as this limit the data to internal access and causes duplication of data storage where both MongoDB and Google's Drive service are utilised. External stakeholders have expressed a clear interest in gaining access to LEO's data where possible and this, amongst other factors, has driven us to explore more tailored platforms to provide increased resource and functionality in the project.

It is important to note that access to LEO's data is subject to the terms and conditions as outlined and agreed upon in the Data Sharing Agreement. This will thus involve a controlled access of data for external stakeholders versus for LEO partners. As some datasets also revolve around ongoing MVS trials or around assets where partners have limited access, these datasets may become available to external stakeholders at a different timeline to internal partners.

3. System Redesign

Recent changes in the reporting of MVS trails has led to a major redesign in how data are collected in LEO. The new data collection form requires a fairly substantial amount information on assets and their use within trials. Thus, many changes (interface and programmatic) were implemented to accommodate this data demand while carefully balancing the ease at which data can be logged online. Furthermore, the present data management tools, though adequate, are not easily scalable, limit access to data, and do not allow for a more holistic data interaction across LEO's various stakeholders. To inject more functionality within LEO's data access and analysis, our team needed to address the issues discussed below and potential solutions are discussed in Section 4.

3.1. **MVS data evolution**

Reporting and the data sharing for MVS trials have now placed greater focus on collecting data specific to the asset at various stages to ensure a more comprehensive evaluation of trials. The first version of the Data Sharing Log therefore required expansion and now allows users to report data in a very flexible manner, using one form for both Background and Foreground data logging. The various steps in conducting MVS trials (known as MVS Procedures) can individually be logged, allowing partners to transition through the form in a customised manner to only log data for procedures associated with their activity (one MVS trial of a flexibility service usually involves multiple LEO partners).

3.2. **Cost-effective data storage solutions**

MongoDB was initially chosen as a solution for LEO's long-term data storage as it offered easy data access through an API and an easily scalable and efficient database. However, the data needs of LEO were underestimated and access to high-resolution data (1-second data files have a substantial size particularly over baselining periods) from substation meters is rapidly increasing as LEO's activities expand. With the frequency of much larger datasets in the order of hundreds of MB to GB becoming more common, MongoDB (an online cloud service) is proving to be a cost-ineffective database. As LEO's data management has replicability at its core, we need to explore options that can be more widely adopted by various stakeholders and *fast-followers* who may not have LEO's full resources within reach.

3.3. Data access and visualisation tools

There is an increasing need, as LEO grows its data streams, to have a one-stop platform that allows for data access, visualisation, and analysis, all under one roof. This platform will open the possibility for partners to interact with LEO's data in a much more agile manner, even providing controlled access for external stakeholders. Having these tools will also increase the accessibility of LEO's data where utility is concerned. Many of the tools developed in LEO for data cleaning and analysis involve strong familiarity with the Python coding software which is not a realistic expectation for the average data user. Thus, user-friendly tools that utilise these cleaning scripts will be developed to account for greater participation within the project.

4. LEO Web Portal[2]

As described in previous sections, current tools are somewhat disconnected and do not allow for a centralised data management system. The following sections will briefly outline the possible solutions currently being explored within the project.

4.1. Key learnings from Data Workshop

A data workshop held by LEO on July 7th, 2020, gave us insight to the internal and external needs for data access. Consistent points were raised with respect to the:

- need for API access to LEO's data for external stakeholders,
- access to LEO's data management and analysis tools,
- and the need to explore the data coming from MVS trials.

Our current system of data management does not easily allow for the adoption of these functions and a more comprehensive platform that can provide these solutions needs to be adopted, one that can also be feasible within other local energy systems.

[2] *This data platform will be separate from, but connected to, the LEO Integrated Land Use Mapping tool.*

4.2. Power BI and Dash

Dash, a backend support for data visualisation and analysis by Plotly, presents LEO with an excellent opportunity for the 'hosting' of bespoke LEO dashboards for its analytical tools. This online and open-access platform (also built on Python) is widely used in many other industries. It will allow LEO the ability to translate complex and code-heavy scripts into user-friendly dashboards that are easily accessible to partners, in and outside of the project.

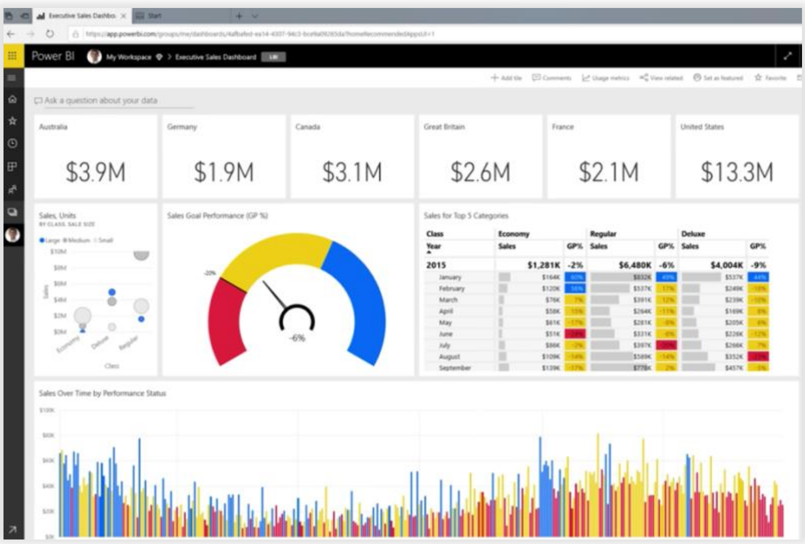
Here are some of the key advantages of using Dash in LEO:

- Allows data to be pulled from the cloud database without data corruption / manipulation.
- Interactive plots of selected data, including provision for spatial mapping.
- Visualisation of data from monitoring equipment (licensing dependent) via API connections with the Eneida portal (contractors of SSEN's monitoring equipment).
- Domain will be hosted through a University of Oxford or private server, in accordance with the Data Sharing Agreement.

If you would like a demo of Dash capabilities, please use the hyperlink above to explore its functionality.

As promising as Dash is for access to analytical tools such as data cleaning, this platform requires a lot of resources for its development, particularly given the data needs in LEO. Power BI, a data analytics platform by Microsoft. This has similar capabilities to the Dash service but provides a much more holistic data management tool for LEO to capitalise upon. The University of Oxford has a licensing agreement with Microsoft which will allow Work Package 4 the ability to use the Power BI platform at no added cost, with the capability to significantly increase storage (also included) that would not be a viable option at a similar storage scale with MongoDB.

Example interface of the Power BI platform (*Source: Microsoft*)



Power BI presents a range of functionality that Dash provides but will require significantly less development time. Comprehensive tools such as cognitive Artificial Intelligence analysis, automatically scaled data tiles and reports, and online sharing of data report via a URL, make this software package very convenient for the average data user.

For a current demo of Power BI's capabilities, have a look at this COVID-19 Dashboard created by Andrzej Leszkiewicz.

These advantages are further complemented by the use of Microsoft Azure for LEO's long-term data storage. These tools will bring many facets of LEO together under the Power BI platform, a tool that can be easily accessed by stakeholders. For instance, it is realistic that a City/County Council can rely on these services with expected in-house resources for long-term data management from similar energy system projects.

However, as Power BI is a closed-access tool developed by Microsoft, customising the platform will be limited as access to the backend of the software is not available. Thus, the Dash platform, though it is not intended that it will be the main hub of LEO's data management, can provide access to tools that allow more customised functionality that Power BI cannot. Case in point, LEO partners are very interested in having an API built to access Eneida data on the SSEN network, and Dash may prove a more useful platform over Power BI in this scenario.

5. Replicability and Next Steps

Data management in LEO needs to keep replicability at its core to ensure that learnings can be effectively translated by *fast-followers* within other local energy systems. Thus, the future design of our system will involve a careful balance of open-access tools, trending and widely adopted software for data analysis, and user-friendly interfaces that increase a project's engagement with its data.

While Power BI is being trailed in LEO as a data solution, LEO will continue to develop tailored tools through Dash that can quickly provide users data functionality that was otherwise limited through LEO's first suite of data tools.