Local Energy Oxfordshire

Sept 2020 | Version One

Data Workshop Report

September 30th, 2019 10.00-16.00

Victoria Grant, Scot Wheeler and Masaō Ashtine

Attendees

Adriano Figueiredo, Ali Sisan, Anitha Sampath, Anne Kersley, David Wallom, Elena Gaura, Ian Dee, Inga Doherty, Jake Verma, Jody Osborn, Kelsey Devine, Laurian Duchamp, Malcolm McCulloch, Martin Taylor, Masaō Ashtine, Matt Blackburn, Paul Goulart, Paul Spencer, Rajat Gupta, Ruth Harris, Sarah James, Scot Wheeler, Seema Dave, Stephen McArthur, Victoria Grant

Summary

The Data Workshop began at 10:00 am in the OeRC Conference Room, with 25 participants attending throughout the day. Engaging discussions were led by Project LEO partners on pertinent concepts of data management, utilization and engagement across a range of audiences. A significant portion of the day's activities was spent in breakout sessions for the **MVS**s (**M**inimum **V**iable **S**ervices) and **KPI**s (**K**ey **P**erformance **I**ndex), where attendees and partners determined key data metrics, needs and management needed for Project LEO. The day closed at 16:00 and saw a successful meeting which has been transcribed into the report that follows.

The following report has been curated by
Victoria Grant, Scot Wheeler and Masaō Ashtine

Overview of Workshop and Goals (*David Wallom, UoO*) - 10:00

Summary points

- A brief overview of the workshop and the agenda was given, stressing that data must be treated as a first-class citizen within the project. Data management is vital to the project and we need to ensure that there is a proper data lifetime in place.

Data Sharing Policy and Plan (*David Wallom, UoO*) - 10:15

Summary points

- Partners are looking at the extension to the collaboration agreement and David went over the standard terminology being used.

- There was a concern with possible limitations introduced in the data agreement and how the data will be shared in the future, both within and outside of the project for future research and impact.

- Parties were asked to complete Schedule B (Part A). Project LEO has already conducted a Data Survey (background data; data that are collected and fed into the project) and this information will eventually make its way into the DSA (Data Service Agreement).

- An important question was raised about what should the partners do about the future licensing of the data for Project LEO. Schedule B (Part B) requires all parties to submit a list of foreground data (generated during the project) that is expected.

- Background data are not expected to be shared wildly but the opposite is true for the foreground data. The foreground data are meant to be shared with the wider PFER program. It does not necessarily have to be open-sourced through Project LEO.

- 'Results' are classed as any data that have been processed whereas the foreground data are in a raw format. 'Results' also encompass the software and any services coming out of Project LEO.

- The data providers are liable to make the data in correct standing and up-to-date, and the receiver's liability is to ensure the data are best managed.

- There will be an online data-sharing log that all parties can access and see what data are being shared in Project LEO. Project LEO wants to understand what data are needed to replicate projects like these and enable learning, and thus the data sharing log is

important in documenting the handling of the data and for the understanding of which data are important or unnecessary.

- EDF raised concerns about the verification process by the partners. Someone can request to see another party's data to ensure compliance but they must understand the handling of the data and the overall objectives. He stated that he can share some of EDF's documentation on this process to help Project LEO.

- The data transfer log will have quite a number of fields as listed below:

    - Unique Transfer ID (if shared)
    - Description of data
    - Type
    - Filename (or file naming convention)
    - Format
    - Data structure
    - Provider
    - Owner (if different to the provider)
    - Receiver
    - Date of creation
    - Date of sharing
    - Documentation details (incl. details on measurement and analysis procedure)
    - Ethical clearance requirements
    - Storage location for provider and receiver
    - Associated license
    - Publication restrictions
    - Access control requirements
    - Date of expiry of data sharing permission or data validity

- The project as a whole will be running its own repository.

- In terms of publication, Project LEO is not intended to place the data into public repositories without special negotiation. Some of the background data may need to be published and thus will involve some level of negotiation.

- The background data survey will filter into the agreement (agreed upon by attendees) and the foreground data registration will come from the MVS form (documentation of what the MVS is going to be - a plan of sorts; agreed upon by attendees) and other related activities.

- All activities that Project LEO does should be considered an MVS.
- There needs to be a confirmation of where the data will be placed for the project.
- All members are encouraged to go back to the survey to reenter any data that they may have missed. David can share what has already been entered to help partners be able to reconsider what they can contribute.

***ACTION**: Review the data with interested partners.*

MOTION**: Malcolm McCullock nominates David Wallom as the Data Manager for Project LEO. Seconded by Anne Kersley. All attendees agree on this nomination.* ***MOTION PASSED

Managing energy data for consumers: (*Jake Verma & Jody Osborn, Catapult*) - 10:45

Summary points

- A brief introduction was given on the definitions of data access and management principles. The importance of common language and vocabulary used in describing the data throughout the project was stressed.
- The Heat Plan Trails project run by Catapult were presented, and the consumer consent needed for data management and collection. Consumer perception of data management and collection were also touched on.
- A brief exercise was conducted to demonstrate some findings from the Heat Plan Trials.
- A second exercise was used to help attendees understand the classification of personal data.

Group paused for a 15' break

LEO Data Survey: Insights (*Rajat Gupta, OBU*) - 11:15

Summary points

- A summary of the Year-1 Data Survey was provided.
- There were 56 total datasets submitted of which 40 were spatial datasets.
- SSEN datasets are still under review and further updates to the Data Survey report will be made.

REMINDER *(to all partners) - The Data Survey link is still live, and partners can update with new datasets when available. Data entered into the survey should capture a list of background info.*

DECISION: *All datasets (even those arising from LEO specifically) should be entered on to the Data Survey.*

- An identifier of background or foreground data will be added to the Data Survey on further iterations of the survey.

FAIR Data Principles (*Masaō Ashtine, UoO*) - 11:45

Summary points

- A summary of what constitutes FAIR data was given.
- **Findable** data - globally accessible and unique. For the data repository, data are given a DOI / appropriate unique identifier but need to be globally unique. Might need to have sub-categories attributed.
- The challenge which all Demonstrators face of attributing globally unique identifiers and the need for developing a local identifier was identified. ERIS said they use bespoke fit

to energy industry which is not aligned to DOIs. EnergyREV noted to perhaps have the development of a 'Technical MVS on Identifier and Catalogue' to understand data across all demonstrators.

- **Accessible -** how do we make data accessible and what identifiers are put in place to make data findable. Protocol to be open, free and universally implementable. But what does this mean for LEO? How do we tailor this to the project taking into account foreground/background data requirements?

- **Interoperable** - need to have an agreed-upon format across partners, which is readable, to ensure data is shareable across partners. Creation of an encyclopedia / agreed vocab and common language of data.

- **Reusable** - accurate, relevant attributes with clear license of usage.

- There was a point raised on cybersecurity of data, and that making data available increases threat factor. Noted the need to make LEO data findable but also secure.

- Need to make sure that LEO data is available to all partners in a format usable to all partners. Recognising that standards will not be prescriptive, but to ensure that all formats of data are used industry-wide.

- Ensuring that data are reusable in LEO will help to increase the impact of the project.

- A possible concern was raised with the scenario of if there's an aggregator and their data is half-hourly but another aggregator has a higher resolution of data, the latter will have a competitive advantage. It becomes worthwhile for the other aggregator to have an increased data resolution. A possible concern of commercial partners.

Review of MVSs and KPIs (Oxford Bus Company PV+ & LEO Land Use Mapping Tool) (*Chairs: Malcolm McCulloch, UoO; Rajat Gupta & Inga Doherty, OCC*) - 12:00

Summary points

- Malcolm presented the model of lean ecosystem transitions giving recognition that there are three groupings of MVSs: Flexibility Services; Geospatial Planning; Informing Policy.

- Development of understanding within Project LEO on who holds risks where certain aspects of the delivery of an asset aren't upheld.

- Point made by ERIS that we need to ensure that LEO KPIs align with the overall PFER performance KPIs.

- Need to make a distinction on what data is needed for MVS data flows required for operation for service / compared to LEO overall KPIs.

MVS1: Flex Services - Malcolm McCulloch

The full workshop group split between MVS1 and MVS2 before and after the lunch break at 13:00

For each MVS, there was a discussion around what data and metadata are needed, what is needed to be stored (and how), and the data users involved. The summary points are listed below for each category.

What data / metadata are needed?

- How are partial wins managed on item 7, MVS1, was received. Level of service to be detailed, rather than just yes/no.

- What is meant by baseline (item 9 MVS1)? What counts as a flex event and how is this verified? Monitoring at a fiscal meter level on flex device itself or collect time and degree heating data or extract info on BMS log. Multiple data sources may be required.

What is needed to be stored and how?

- Still to be defined

Data users - when do they receive data, how, where?

- Marketplace participants to start with.

Other points raised

- There was a discussion around the need to audit the data that are required before, during and after the flex event so that it is possible to critically playback the decision

events during the MVS. This will inform investment decisions. Packaging data for the particular MVS to replay scenario - so that you can see flex event requirement, how market placed responded etc.

- How to impart a process to pay partners/stakeholders who input data onto a tool but also to extract value from this? Look at stock market data availability.

MVS2: Land Use Mapping - Inga Doherty & Rajat Gupta

- There was a question of how to manage increases in the demand of energy attributed to the increase in housing, EVs and heat within Oxfordshire.

- A question was raised of how detailed we want the land use map to be. Does it go down to individual land parcel information?

What data / metadata are needed?

- Group came up with a list of datasets that are desirable:
    - Add weather information on to list - i.e. Met. Office, Historical Information from the ECMWF.
    - River heights - to understand if there's capacity for hydro etc.
    - Include actual current demand and predicted demand (SSEN).
    - Add in CORINE Land Cover data.
    - How can we incorporate transport data into the map.
    - ONS data - indices of deprivation.

- ERIS has a similar data tool which uses similar datasets which could be made available to LEO land use mapping tool.

- A question was raised on how to link spatial data with smart metering data. ERIS know of datasets available relating to this and which ones are paid.

ACTION: *ERIS to pass on a contact of where these datasets can be obtained.*

- There was an idea that the map is one part of the tool, but that there should be descriptions added to outputs to make the tool more usable and understandable to users.

- A question was raised on how to build in a 'health check' of datasets and layers. Need to ensure that ground-truthing / validation testing of data used within the tool is undertaken.

- Need to the map to be embedded inside metadata and description input so that the receivers of the tool can interpret data in the correct manner.

- Need to identify who is using the mapping tool, how many times, and for what purpose. There is a need to determine how reports are being generated and how these are being used.

What is needed to be stored and how?

- Project LEO should utilize a 'metadata lake' of sorts for storing the metadata.

- Partners had a discussion around potential partnerships: DAPHNI? Catapult?

Data users - when do they receive data, how, where*?*

- There was a question on are certain types of data in the tool given different access so that certain users can view? Is it a general public tool or just kept for planners/developers etc.?

- There is a need for the land use mapping tool to have its own data-sharing agreement between partners involved as this will go beyond LEO.

- An idea was raised that an output of the mapping tool could be published but this would not provide raw data, i.e. data can't be traced back.

- Need a tiered approach to making data available to various users.

    - Public users - what access do they need/want? Not data contributors.
    - Developers / Housing Associations.
    - Planners within County Council / Consultants?

- Use-case analysis needed for the mapping tool. Legacy of how this tool is managed after the lifetime of LEO. What happens if LEO datasets are used? Are they still accessible after the 3 years? Are they kept up to date? In terms of sponsorship of the platform, is it something that the LCH can help fund?

- A business model activity discussion needed. PAYG or generation of reports. Need to develop use-cases. How can LEO commercially utilise this tool?

- Discussion needed on ownership and replicability but the need to ensure that it sits within the Local Authority remit.

KPI Dashboard (*Scot Wheeler, UoO*) - 14:45

Summary points

- Definition of Project LEO's Oxfordshire boundary is SSEN area and County Council.

There was a breakout session for all workshop participants to assess the needs for the KPIs for Project LEO that were previously established. Participants were also able to lend advice on removing/adding existing and new KPIs. The following tables give a summary of the points:

KPI	***Raw Data***	***Resolution***	***Responsible Partner***	***Update Frequency***
Dispatchability *(availability)*	*Piclo Database* *(Geographic Location + Network Location)*	*Per Device*	*Piclo*	*Quarterly*
Number of Systems Participants *(commercial + non-commercial users; non-participants + passive participants)*			*LCH / Piclo Aggregators (Nuvve, EDF)*	
Network utilization *(number of times overutilization threshold)*	*Substation metering* *Smart Meter Data* *Asset Network / Rating*	*30 minutes or 6 minutes w.r.t MVSs* *Weekly* *Per Feeder*	*SSEN* *LCH*	
Project Engagement	*Website hits* *(who, why)* *Newsletter subscribers + responses* *Citations*		*SSEN*	*Monthly*

Demand, CO_2 Intensity	*kWh, kW (Peak Power Demand), $kgCO_2$ / kWh*	*Half-hour data capture*	*DNO*	*Quarterly*
Renewable Generation (decentralized generation)	*kWh*	*Half-hourly*	*DNO*	*Quarterly*
Change in Self-Sufficiency	*kWh*	*Annually*	*DNO*	*Quarterly*
*Commercially Verified Services**				
Energy Equity	*Difference between max. and min. energy*			
Energy Affordability	*End-user cost / kWh*			
*Diversity of Technology Owners**				
*New Technology**				
End-of-LEO KPIs				
*Resiliency & Sustainability of MVSs**				
*Efficiency**				
*Exploitation**				
Job Creation	*Jobs per partner + wider 'approved' suppliers*	*Full-time equivalent (FTE)*	*County Council*	

** KPI details TBD*

Data Value (*David Wallom, UoO*) - 15:40

Summary points

- There needs to be a balance of how much of the data will be re-used / against the cost of re-generating data.

- Need to understand the lifetime of the data required for each of the KPIs. When is it not needed to be stored?

- Recognition was given to the importance of data catalogue to canvas input and interest from industry.

- The question was raised of how to share enough data publicly but also to ensure that data is protected where necessary, to ensure commercial viability. There needs to be a balance between the openness of data but also keeping data close to LEO.