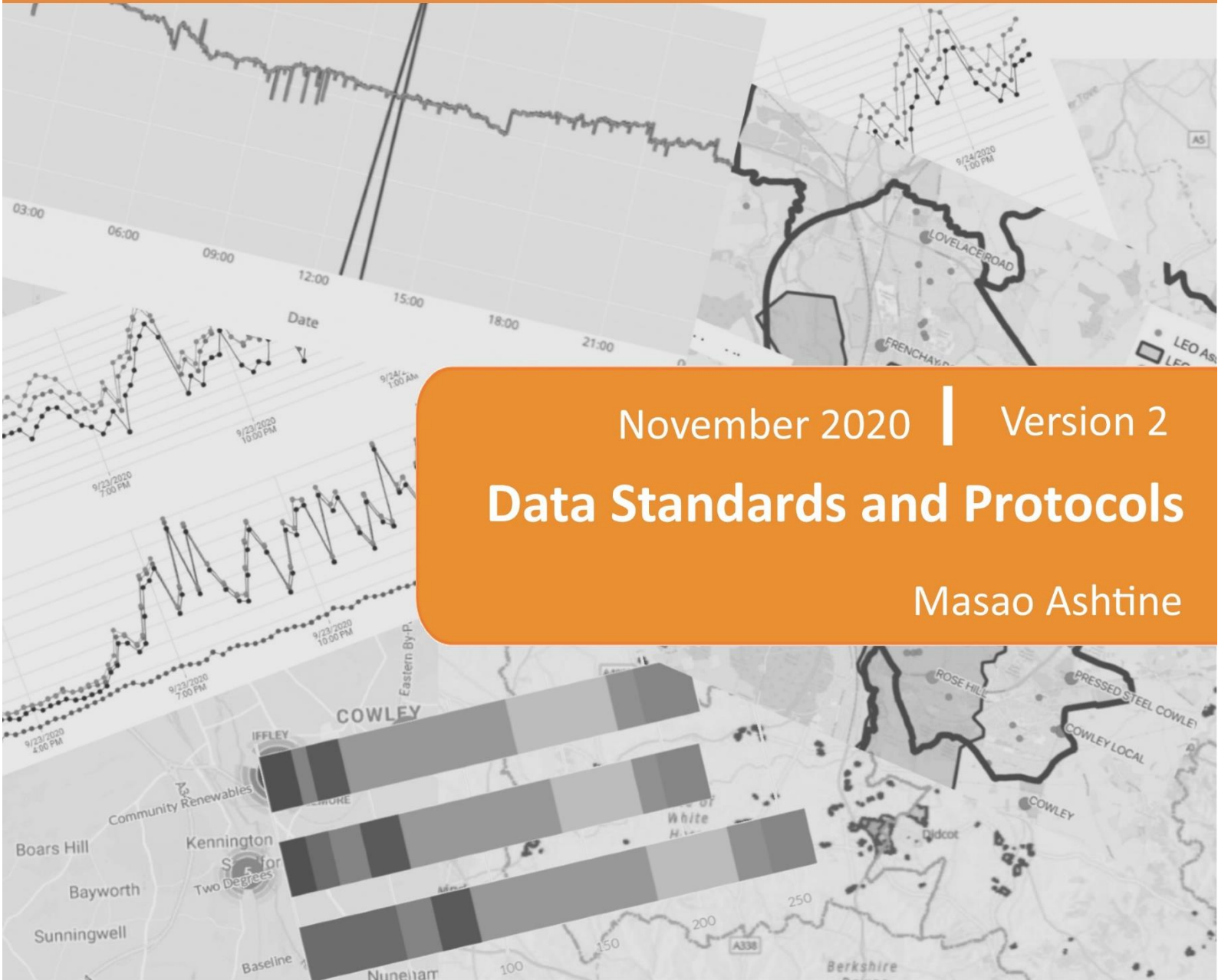




Local Energy **Oxfordshire**



November 2020 | Version 2
Data Standards and Protocols
Masao Ashtine



Report Title:	Data Standards and Protocols
Author(s):	Masaō Ashtine
Organisation(s):	University of Oxford

Version:	2.0	Date:	22/10/2020
Workpack*:	WP4	Deliverable:	4.13
Reviewed by:	Rajat Gupta		
Date:	27/10/2020		
Signed off by:	David Wallom		
Date:	16/11/2020		

Can be shared (Y/N):	Internally		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	4
2.	Definitions and Interpretation	4
3.	Project LEO Data Manager	5
4.	Project LEO Party Data Coordinators	5
5.	Vision and Principles	6
6.	Good Data Practices	7
6.1.	<i>Data Principles</i>	7
6.2.	<i>Data Formats</i>	9
6.3.	<i>Data Translators</i>	9
7.	Recent developments since last publication	10
7.1.	<i>Data Certificates</i>	10
7.2.	<i>Standardising datasets</i>	10

1. Executive Summary

This report outlines the main processes and standards involved in the management of shared data (and associated metadata) within Project LEO. The report is an update of the 2019 report where more information has been added on processing and quality control of Foreground data generated in Project LEO. The report defines the main terms used to express data management in LEO and lists Data Coordinators of the project and the standards for data management covering pre-processing to post-processing and analysis. The report concludes with a description of current considerations of the standards to which data are upheld.

Note, this report does not outline the detailed steps used to handle data in LEO, but the standards which data management is expected to meet.

2. Definitions and Interpretation

The following words and phrases, when used in this document, shall have the meanings given to them below:

- 2.1. **Background Data** - The data or information that is generated outside of Project LEO and is more particularly described in Schedule 2 Part A to this Data Sharing Agreement which is disclosed by the Provider to the Receiver. For the avoidance of doubt, Results and Foreground Data do not form part of the Background Data.
- 2.2. **Confidential Information** - Any private, secret or commercially sensitive information which is disclosed by one party to the other in connection with Project LEO and is marked as confidential by the disclosing party at the time of disclosure or is clearly confidential by its nature.
- 2.3. **Data** - Is used as a collective for Background Data and Foreground Data as defined in Clause 1. "Definitions and Interpretation" of the Data Sharing Agreement.
- 2.4. **Data Sharing Log** - A register of all data sets shared between project partners with associated metadata, maintained by the Project LEO Data Manager.
- 2.5. **Employee** - A person who is contracted by one of the LEO partners to work on Project LEO.
- 2.6. **Foreground Data** - Any data or information which is obtained or generated in the course of performing Project LEO and is more particularly described in Schedule 2 Part B to the Data Sharing Agreement which is disclosed by the Provider to the Receiver. For the avoidance of doubt, Background Data and Results do not form part

of the Foreground Data.

- 2.7. **Party Data Coordinator** - The person appointed by each Party who is responsible for ensuring compliance with the Data Sharing Agreement and liaising with the Project LEO Data Manager.
- 2.8. **Personal Data** - Has the same meaning as defined in Clause 1.1 of the Data Sharing Agreement.
- 2.9. **Project Data Manager** - The person appointed by Project LEO as the first point of contact for Project-wide data, and who is responsible for curating the Project LEO Data Sharing Log.
- 2.10. **Provider** - Any Party that discloses the Background Data and/or Foreground Data to the other Parties under the Data Sharing Agreement.
- 2.11. **Purpose** - The use and analysis of the Background Data and Foreground Data for the purposes of Project LEO and the use of the Results for further research and academic purposes including publication.
- 2.12. **Receiver** - Any Party or each Party receiving the Background Data and/or Foreground Data from the Provider under the Data Sharing Agreement.
- 2.13. **Results** - Any data or information derived from the analysis of the Background Data or Foreground Data in the course of Project LEO.

3. **Project LEO Data Manager**

Within the Project LEO Data Workshop held on September 30th, 2019, a motion was passed and subsequently approved to have Professor David Wallom of the University of Oxford act as the overall Data Manager for Project LEO.

4. **Project LEO Party Data Coordinators**

The following people will act as the main coordinators and data managers within their respective organisations. These partners represent the officially listed Parties within the Data Sharing Agreement.

Southern Electric Power Distribution Plc.

Andrew Waterston; Kyle Murchie
andrew.waterston@sse.com kyle.c.murchie@sse.com

The Low Carbon Hub

Rachel Boyes-Watson
rachel@lowcarbonhub.org

Oxford Brookes University

Rajat Gupta
rgupta@brookes.ac.uk

EDF Energy R&D UK Ltd

Benjamin Mousseau
benjamin.mousseau@edfenergy.com

Oxford City Council

Ruth Harris
rharris@oxford.gov.uk

Oxfordshire County Council

Anitha Sampath
anitha.sampath@oxfordshire.gov.uk

Nuvve Limited

Paige Mullen
paige@nuvve.com

Open Utility Ltd

Kelsey Devine
kelsey@piclo.energy

The University of Oxford

Masaō Ashtine
masao.ashtine@oerc.ox.ac.uk

5. Vision and Principles

To manage data effectively within and beyond the project's lifetime to ensure that all associated partners and authorised stakeholders have shared access to Foreground and Background data which are stored and managed under the FAIR (Findable, Accessible, Interoperable, Reusable) guiding principles. Tools such as [FAIRshake](#) can be used as a measure of data readiness and accessibility within Project LEO.

Results of the project will also fall under this vision to ensure that data analysis is effectively communicated and disseminated amongst all Parties concerned. All data will be managed in accordance with the Good Data Practices outlined within this document, and the Data Sharing Agreement agreed-upon by all Project LEO partners. This document will continue within the project as a living document and form the basis of documenting data management within Project LEO. As such new major versions will encompass significant additions, for example as a result of a new MVS (Minimum Viable Service) and/or any new learnings from their start-up or operation.

6. Good data practices

6.1. Data principles

The following sections outline the main principles which govern good and FAIR data practices within Project LEO.

- 6.1.1. Foreground Data must be generated using sound scientific¹ techniques and processes. Foreground Data must also be accurately recorded in accordance with good scientific practices overseen by an employee.
- 6.1.2. Data must be analysed and used appropriately, without bias and in accordance with good scientific practices. Data and the results must also be stored securely, appropriate to support the requirements of any relevant licences, institutional data policy or current UK Data Protection Legislation, and be easily retrievable.
- 6.1.3. Appropriate levels of confidentiality and compliance with GDPR must be maintained during collection, use, transmission, storage and destruction of data associated with Project LEO.
- 6.1.4. The integrity (accuracy and completeness), of all data, must also be maintained during any sharing and/or storage of data.
- 6.1.5. Data provenance records must be kept to allow an employee to easily demonstrate project activities and reconstruct key decisions made during the conduct of the research, presentations made about the research, and conclusions reached in respect of the research.
- 6.1.6. Each Party must have the right, on no less than 30 days' written notice, to visit any other Party to verify that it is complying with the above practices and procedures. It is important however that the visiting Party must understand the handling of the data and the overall objectives.
- 6.1.7. Project LEO, through the nominated Project Data Manager, will record all instances of data sharing in a Data Sharing Log such that there is awareness of the transfer of data between Parties. The Data Sharing Log will be available to all Parties via the Project LEO SharePoint.

¹ The systematic collection and processing of data related to all project activities and research which follows structured and widely acceptable methods of research and analysis.

- 6.1.8. It is the responsibility of the Party Data Coordinators to ensure that the Project LEO Data Manager is made aware of any data transfer between Parties and provide the associated metadata.
- 6.1.9. For all data within Project LEO, if it is either (a) being shared between Parties or (b) generated as part of Project LEO, the following information (listed in no particular order below) are recorded, plus any further reasonable information as requested by the Project LEO Data Manager:

<i>Metadata Field*</i>	<i>Further notes</i>
Unique transfer ID if shared	<i>All datasets are marked with a unique LEO ID in the form of LEOD{Dataset number: 3 digits}F{File number: 2 digits} for Foreground and Background Data e.g.: LEOD023F01. This ID is embedded into a datasets filename as seen below.</i>
Description of data	<i>Basic and easily understandable description of the data. This is captured within the Project LEO Data Log.</i>
Type	<i>Reports standard set of data types (Temporal, Spatial etc.)</i>
Filename (or file naming convention)	<i>All datasets are marked with a unique LEO ID in the form of {MVS ID}_{Procedure Step}_{Data Provider}_{Date Shared}_LEOD{Dataset number: 3 digits}F{File number: 2 digits} for Foreground Data e.g.: MVSA1_1_1_P2_LCH_25022020_LEOD023F01. Background Data will take the form of BKGRD_{Data Provider}_{Date Shared}_LEOD{Dataset number: 3 digits}F{File number: 2 digits} e.g.: BKGRD_UoO_25022020_LEOD000F01</i>
Format	<i>Reports standard set of data formats (CSV, Shapefile etc.)</i>
Provider Owner (if different to the Provider) Receiver Date of creation Date of sharing	<i>Dates recorded in the YYYYMMDD e.g. 20191031</i>
Documentation details (including details on measurement or analysis procedure)	<i>Raw and analysed data must be accompanied by a brief description of their collection and/or methodologies, or a reference citing any external documentation giving details.</i>
Ethical clearance requirements	

cont'd

Storage location for provider and receiver

Clearly defined URL / API / Location for data storage

Associated Licence

Publication restrictions

Date of expiry of data sharing permission or data validity

** Please note that MVS data are shared with more metadata fields than listed to aid in the understanding of process maturity and asset specifications. These are not included here as they do not represent the minimum level of metadata gathering to meet outlined data standards.*

Within Project LEO, certain metadata fields will be standardised to ensure data consistency and accurate provenance (as detailed above). The table contains important metadata fields that will allow for good practices of data management in Project LEO, however, this list is not the complete list of detailed standards and formats that will be controlled by the Project Data Manager.

As this document further develops and matures through later revisions, it will undergo periodic review and ratification by all of the Party Data Coordinators.

6.2. Data Formats

As Project LEO develops, the data collected within the LEO data system will initially be stored in a native, as delivered, format. Investigation will continue as to the utilisation of a more universal format (if and where possible) for both temporal and spatial data. The main objective for the standards set forth in the document with regards to data formats is to ensure that all data within Project LEO satisfy the objectives of the principles discussed in [Section 5.1](#).

6.3. Data Translators

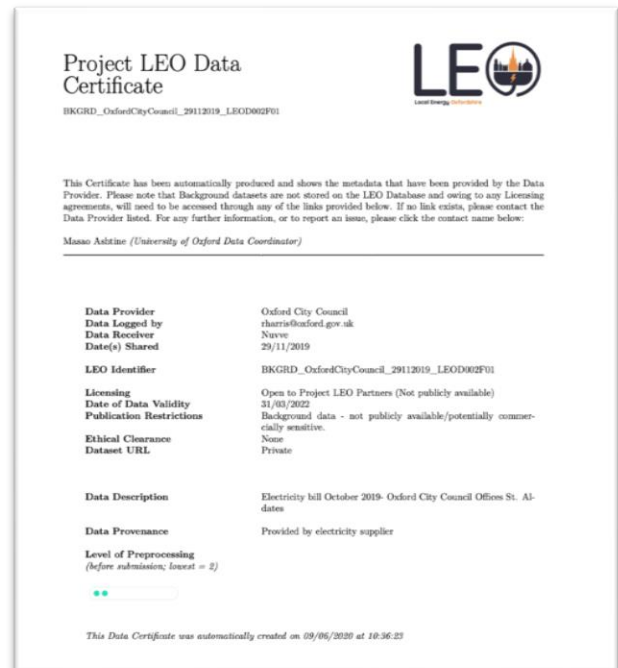
As with data formats, Project LEO will define and use appropriate data translation software and/or procedures to ensure that data from one format are properly converted to another in keeping with the data principles and practices defined within this document.

7. Recent developments since last publication

7.1. Data Certificates

Since the last version of this report in 2019, Project LEO has adopted a system whereby all shared datasets (each file falling within a dataset to be specific) are associated with a Data Certificate. This document is automatically produced based on data entered when datasets are shared by project partners and gives a summary of the key metadata associated with the relevant file.

These certificates increase the data provenance within the project and allow partners and relevant stakeholders to quickly understand the data being shared.



7.2. Standardising datasets

With the increased activity of Project LEO and associated connected assets and plug-in projects, datasets will invariably possess an increasing amount of format diversity. CSV (comma-separated values) is the most common temporal dataset format within LEO. Owing to the various software and means of data processing before data are shared, partners are able to share data with a wide variety of structure that can exacerbate a number of possible bad practices/faults that impact on data usability (headerless data columns, excessive empty rows of data, missing data and bad formatting etc.). Standardising spatial data is not of primary concern owing to the inherent flexibility of such datasets, but the same metadata standards are to be maintained.

Thus, within the project, we are considering tools that allow partners the opportunity to run an automated 'health scan' of their datasets, to ensure that further processing, cleaning and management of data can be maintained at a high quality that meets the outlined standards and protocols listed in previous sections. Further information on how data are handled after being shared can be found in the Data Cleaning and Processing report for the project.