

BD4OPEM

Big Data for OPen innovation Energy Marketplace

Deliverable 1.2

Data Management Policy

Document information

Deliverable nr	D1.2
Deliverable Name	Data Management Policy
Version	1.0
Release date	30/06/2020
Dissemination level	Public
Status	Submitted
Author	Nigel Claridge, Yasmina Ganse (SUST)



Document history:

Version	Date of issue	Content and changes	Edited by
0.1	18/05/2020	First version	Sustainable Innovation
0.2	01/06/2020	Second version	Sustainable Innovation
0.3	05/06/2020	Third version	Sustainable Innovation
0.4	22/06/2020	Fourth version. Updates implemented after peer-review process	Sustainable Innovation
1.0	29/06/2020	Fifth version. Updates implemented after GA review process	Sustainable Innovation

Peer reviewed by:

Partner	Reviewer
ICOM	Isodoros Kokos
WEP	Amit Eytan
UPC	Ismael Bravo
UPC	Mònica Aragüés
ODT	Albert Gonzalez

Deliverable beneficiaries:

WP/ task
All

Table of contents

1	Executive summary	6
2	Introduction	7
3	Data Summary	9
	3.1 Purpose of data collection	9
	3.2 Types and forms of data that will be collected	10
	3.2.1 Danish pilot	10
	3.2.2 Turkish pilot	11
	3.2.3 Belgian pilot	12
	3.2.4 Spanish pilot	12
	3.2.5 Slovenian pilot	12
	3.3 Five stage management approach	13
4	Fair Data	14
	4.1 Making data Findable, including provisions for metadata	14
	4.2 Making (pilot) data openly Accessible	15
	4.3 Making data Interoperable	15
	4.4 Increase data Re-use	16
5	Allocation of resources	17
6	Data security	18
	6.1 GDPR	18
7	Ethical aspects	19
8	Other issues	20
9	Conclusions	21
10	References	22

Abbreviations and Acronyms

Acronym	Description
CAg	Consortium Agreement
GAg	Grant Agreement
DoA	Description of Action
EC	European Commission
GA	General Assembly
CG	Core Group
PC	Project Coordinator
PO	Project Officer
QM	Quality Manager
FM	Financial Manager
TC	Technical Coordinator
EDM	Ethics & Data Manager
DCM	Dissemination & Communication Manager
IEM	Innovation & Exploitation Manager
SAB	Security Advisory Board
AB4DE	Advisory Board for Dissemination and Exploitation
TL	Task Leader
ToC	Table of Content
WP	Work Package
WPL	Work Package Leader

List of Figures

Figure 1. Location of the BD4OPEM pilot sites 10

1 Executive summary

The project recognises the importance of regulating data management issues. A solution will be considered where consortium partners will deposit all underlying information on data-related business processes (data storage, data provisioning, processing etc.) of the BD4OPEM solution clearly and transparently.

The purpose of the Data Management Policy (DMP) is to identify the main data management elements that apply to the BD4OPEM project and the consortium. This document is the first version of the DMP and will be reviewed as soon as there is a clearer understanding of the types of data that will be collected.

Given the wide range of sources from which data that will be collected or become available within the project, this document outlines that the consortium partners will consider embracing and applying the Guidelines on FAIR Data Management in Horizon 2020; “In general terms your data should be ‘FAIR’, that is Findable, Accessible, Interoperable and Re-usable” [1], as information about data to be collected becomes clearer.

Therefore, as described in this document, the consortium will deposit the public data produced within or collected for the purposes of the project in an open data repository once the exploitation rights are safeguarded; permit the user to access, mine, exploit, reproduce and disseminate free of charge; provide information about tools and instruments necessary to the project results whenever possible.

As clarified in the DMP, An Ethics & Data Manager (EDM), SUST, will manage the compliance of all project activities with good practices (legal, ethical, privacy and data protection) from the beginning of the project ensuring that data gathering procedures conform to the Regulation (EU) 2016/679 of the European Parliament (GDPR) and of the Council of 27 April 2016 “The protection of natural persons in relation to the processing of personal data is a fundamental right” [2]. Each partner shall ensure compliance with any national laws or regulations. A Security Advisory Board (SAB) will be established to ensure that the measures defined in WP5 are followed and complied with, preventing data from ending up in the hands of unauthorised external entities.

As Data Security and Data Privacy are of particular concern, BD4OPEM will consider the GDPR privacy principles from conception to design in order to build a novel privacy-preserving architecture.

2 Introduction

The purpose of the Data Management Policy (DMP) is to identify the main data management elements that apply to the BD4OPEM project and the consortium, in terms of project research data and internal repositories.

The DMP covers the complete research data life cycle. It outlines the types of research data that will be generated or collected during the project, the standards that will be used, how the research data will be preserved and what parts of the datasets will be shared for verification or re-use.

The DMP is a living document, which will evolve during the lifespan of the project, particularly whenever significant changes arise such as dataset updates or changes in Consortium policies. This document is the first version of the DMP, delivered in Month 6 of the project and will be reviewed as soon as there is a clearer understanding of the types of data that will be collected.

It includes an overview of the datasets to be produced by the project, and the specific conditions that are attached to them. Although this report covers a broad range of aspects related to the BD4OPEM data management, upcoming versions will provide more detail on particular issues, such as data interoperability and practical data management procedures, take form and are implemented by the project consortium.

The core of this document is divided into five sections after the introduction to cover all issues related to the DMP:

- [Data summary](#) (Section 3)
- [Fair data](#) (Section 4)
- [Allocation of resources](#) (Section 5)
- [Data security](#) (Section 6)
- [Ethical aspects](#) (Section 7)
- [Other issues](#) (Section 8)

This DMP describes the data management life cycle for the data to be collected, processed and/or generated. As part of making research data findable, accessible, interoperable and reusable (FAIR), the DMP will include information on how BD4OPEM will address and adapt to the principles regarding:

- Handling of research data during & after the end of the project
- What data will be collected, processed and/or generated
- Which methodology & standards will be applied
- Whether data will be shared/made open access
- How data will be curated & preserved (including after the end of the project)

The DMP will be updated over the course of the project whenever significant changes arise, such as (but not limited to):

- New data
- Changes in consortium policies (e.g. new innovation potential, decision to apply for a patent)
- Changes in consortium composition and external factors (e.g. new consortium members joining or old members leaving)

3 Data Summary

3.1 Purpose of data collection

BD4OPEM stands for Big Data for OPen innovation Energy Marketplace. Global demand for electricity is increasing, and energy systems have moved from an analogue to an interconnected real-time digital world. Huge amounts of data, mostly unused or underused, are available offering great potential to develop exciting new services. These services will improve the planning, monitoring, operation and maintenance of electrical distribution systems, all available from an open innovation marketplace.

A seamless link between energy stakeholders and solutions developed will ensure secure data flows between data providers and solution providers, resulting in new data-driven business models, enhanced asset management and consumer participation in energy balancing. Target user groups will be able to find relevant solutions provided by different specialised companies.

BD4OPEM objective is to develop an Analytic Toolbox to interconnect data providers and service providers and to demonstrate, validate and verify the feasibility of the services produced in different contexts and countries prior to any deployment in a trans-European context.

Five pilot sites (Spain, Turkey, Slovenia, Belgium and Denmark), who provide the initial input data, will demonstrate the usefulness and the usability of the services developed. BD4OPEM will collaborate with these pilot sites from the European energy sector, who will play a central and essential role in the project. Not only will they provide the project with massive amounts of data to enable the creation of all BD4OPEM services, but they will also act as testbeds for testing and validating the services as they are deployed.

Eighteen innovative services to enhance Operation and Maintenance, Monitoring and Planning decision-making will be tried and tested in five pilot sites, all available from the BD4OPEM Open Energy Innovation Marketplace.

3.2 Types and forms of data that will be collected

The following is a high-level summary of the data to be collected. For more detailed information, please see document D3.1, which gives a fully detailed account of all the data to be collected at each pilot site.

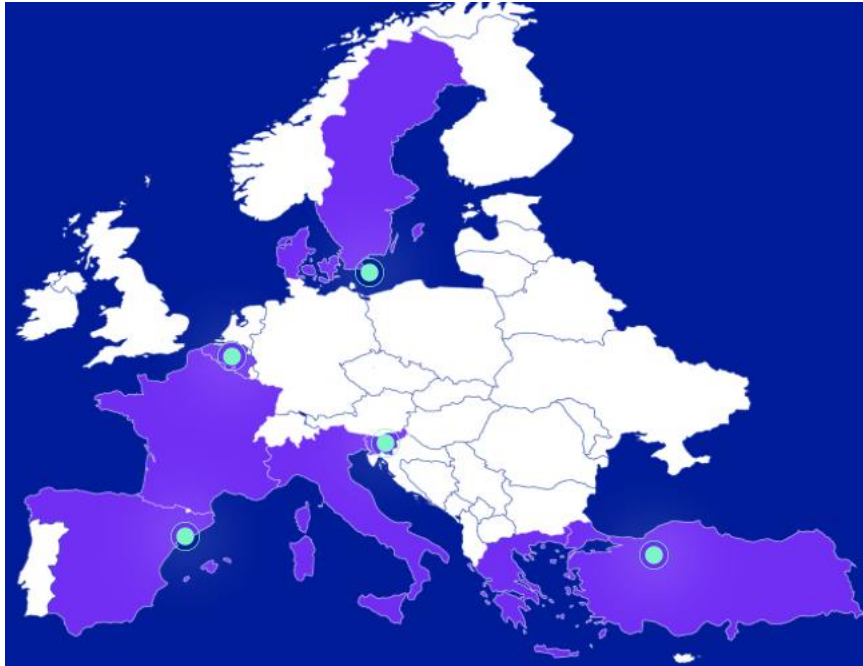


Figure 1. Location of the BD4OPEM pilot sites

3.2.1 Danish pilot

As a customer of the local DSO, NUV (Nuvve) operates some 30 bidirectional 10 kW chargers on the Danish island of Bornholm. This represents a scaled model of the Danish renewable integrated power system operating in grid-connected and island mode.

The local DSO Bornholms Energi & Forsyning (BEOF) supplies electricity to more than 28000 customers including the app. 30 EVSE's of Nuvve. BEOF accepted to share relevant grid operational data as well as anonymized data from smart meters. Part of the Pilot is to develop Services for analysis and comparison of grid topologies and customer loadings, with EV penetration, driving and charging patterns, and of temporal load flows will be investigated to estimate flexible provision for the power system and its market, that can maximize the benefits of the system, aggregator and users. Bornholm is a well-advanced energy island owning and running a state-off the

art microgrid. It includes a thermal and electricity grid, waste water recovery, a high-speed glass-fibre telecom network and a total of 33 HV transformers divided over HV 18 substations. Energy production and storage include photovoltaics (817 kWp), CHP 2.8MW, and 3 emergency generators (5.25 MVA), and a total capacity of 2,5 MWh in battery storage, mainly under the form of UPS. The local EV charging infrastructure is expanding besides the NUVVE V2G EVSE's to include more uni-directional chargers and fast chargers. The financial bookkeeping and billing to the different consumers in the microgrid is carried out by means of ERBIS software platform. Cutting edge control technology and maximal reliability are the focus points of this demonstration site.

Source of data provided by Nuvve:

- EV data
- EV charging infrastructure data

3.2.2 Turkish pilot

The pilot site in Turkey has differential characteristics respect the other DSOs in Slovenia and Spain, both with a huge smart meter's deployment. In this pilot, there are 3 TSO substation and 9 High Voltage – Medium Voltage transformers. This region is consisting of urban and rural areas so that these 3 TSO substations are feeding a variety of customers with more than 1000 distribution transformers whereas, connected with 70 MW installed distributed generations which are connected from medium voltage. In this area, the PV penetration is about 50% within this 70 MW installed capacity. Because of the high penetration of PV, there are some power quality problems, such as overvoltage, voltage fluctuation, congestion and reactive power compensation issues are faced.

The region consists of urban and rural areas where the 3 TSO substation and 9 High Voltage Medium Voltage transformers supply a variety of customers. In this area, the PV penetration is very high. Daily production and consumption are constantly changing, and meeting supply and demand is a challenge.

Source of Data provided by OEDAS:

- Topology & Observability of MV network
- SCADA, GIS, Analysers
- Energy storage infrastructure
- PV, Hydro, Geothermal, wind generation

3.2.3 Belgian pilot

As a customer of the local DSO, the Brussels Health Campus is a well-advanced energy island owning and running a micro-grid that is able to operate in “island” mode for five consecutive days. The hospital and part of VUB is a critical environment where grid security is highly prioritised.

Source of data provided by VUB:

- Digital meters
- Topology & Observability of LV network
- Client clustering and consumer profiles (on aggregated substation or building level)
- SCADA, Analysers
- Generation: PV (potentially genset and CHP)
- Rechargeable Energy storage systems
- Electric vehicle charging profiles

3.2.4 Spanish pilot

Estabanell is both a Distribution Systems Operator (DSO) and retailer. The main business activity is electricity distribution. With a network of over 1.100 km, it supplies more than 56.000 power points, with two substations where it connects to the transmission network at 220kV, distributing electricity through more than 800 secondary substations. Generating sources include PV and micro-hydro.

Source of data provided by Estabanell:

- Smart meters (55,400)
- Topology & Observability of MV network
- SCADA, GIS, Analysers
- Consumer profiles (patterns & consumption)
- PV generation
- Large scale battery
- EV charging infrastructure data

3.2.5 Slovenian pilot

Elektro Celje is one of five DSOs in Slovenia, covering 22% of the territory. Its electricity infrastructure is extensive, and it supervises, manages and operates the

electricity distribution network supplying over 170.000 customers of which 125.800 (75%) are equipped with smart meters.

Source of data provided by Elektro Celje:

- Smart meters
- Topology & Observability of LV network
- Client clustering and consumer profiles (patterns & consumption)
- SCADA, GIS, Analysers
- PV, Hydro generation

3.3 Five stage management approach

The BD4OPEM project plans to follow a five-stage management approach that outlines the process for data handling, also considering FAIR data principles described in section 4:

- Data acquisition – Data (see 3.2) will be received from our pilot sites and external entities (e.g. TSO transparency platforms). Data acquisition will be compliant with state-of-the-art security protocols.
- Data storage - Data will be stored in the BD4OPEM solution with special attendance on privacy and security issues of sensitive data (e.g. through encryption). Apache Hadoop is one solution discussed that will facilitate the storage and management of the data. The storage component is called HDFS (Hadoop Distributed File System).
- Data protection - We plan to limit access to our database servers by using a firewall. We plan to manage access to servers by creating varying levels of authorisation within our team. We authorise only the senior technical staff for full access. Finally, database, server and user passwords are automatically changed regularly.
- Data retention - We plan to replicate the data in more than one database, using master and slave architecture (i.e., master server replicates itself onto slave servers). We are considering replicating the server sets, i.e. Sharding (Sharding is a type of database partitioning that divides extensive data into manageable smaller and faster data sets called shards). Even if some of the servers do go down, the project will never lose data, and the BD4OPEM system will not go down. Also, we back up our data regularly to avoid any loss of data.

- Data destruction – A plan for handling the data after the end of the project will be produced as we move towards the end of it. The exploitation of the project result will determine the scope and content of the plan.

4 Fair Data

In order to ensure a soundly managed data, BD4OPEM will consider how we embrace and apply the FAIR principles, to make data Findable, Accessible, Interoperable and Reusable. All partners in BD4OPEM will adhere to the FAIR principles as they manage data.

The identifiers below for the Fair data principles are described in Force 11:

<https://www.force11.org/group/fairgroup/fairprinciples>

4.1 Making data Findable, including provisions for metadata

To be Findable:

- F1. (Meta)data are assigned a globally unique and eternally persistent identifier.
- F2. Data are described with rich metadata.
- F3. (Meta)data are registered or indexed in a searchable resource.
- F4. Metadata specify the data identifier.

BD4OPEM plans to establish whether the data produced and/or used in the project are discoverable with metadata, identifiable and traceable through a standard identification mechanism (e.g. persistent and unique identifiers such as Digital Object Identifiers). This will include naming conventions and identifying search keywords. Also, what metadata will be created will be considered and, in the event, that case metadata standards do not exist, we will outline the type of metadata created and how it will be created.

4.2 Making (pilot) data openly Accessible

To be Accessible:

- A.1. (meta)data are retrievable by their identifier using a standardised communications protocol.
 - A1.1. The protocol is open, free, and universally implementable.
 - A1.2. The protocol allows for an authentication and authorisation procedure, where necessary.
- A2. Metadata are accessible, even when the data are no longer available.

BD4OPEM plan to establish where we will store and share the data, which data repository will be used and why? This will become evident once we have a better understanding of the data to be collected, but one possible solution is Zenodo.

All data produced and/or used in the project will be made openly available within the project. Pilot data which are marked open access can be found in D3.1. From external data, market data are open access. If certain datasets cannot be shared (or need to be shared under restrictions), explanations will be provided based on legal and contractual reasons. During subsequent versions of this document, the appropriate intellectual property protection will be described.

4.3 Making data Interoperable

To be Interoperable:

- I1. (Meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.
- I2. (Meta)data use vocabularies that follow FAIR principles.
- I3. (Meta)data include qualified references to other (meta)data.

The aim is to follow relevant international standards, e.g. SAREF, CIM, which makes achieving interoperability possible.

For the next update of the DMP, the project will identify data and metadata vocabularies, standards or methodologies to be followed in order to make our data interoperable. Where possible, we plan to follow standard vocabularies for all data types present in our data set, thus allowing inter-disciplinary interoperability.

Our aim is to make the solution as open as possible, and we should, therefore, set out any necessary ontologies of this kind. This is a decision that has not yet been taken.

4.4 Increase data Re-use

To be Reusable:

- R1. Meta (data) have a plurality of accurate and relevant attributes.
 - R1.1. (meta)data are released with a clear and accessible data usage license.
 - R1.2. (meta)data are associated with their provenance.
 - R1.3. (meta)data meet domain-relevant community standards.

Data re-use policies will be outlined, discussed and decided as the project moves forward and presented in greater detail at the next revision. Concerning licenses related to the dataset, the project will consider Attribution-Non-Commercial 4.0 International (CC BY-NC 4.0) licence when it comes to increasing the possibility of re-use.

5 Allocation of resources

An Ethics & Data Manager (EDM), SUST, will manage the compliance of all project activities with good practices (legal, ethical, privacy and data protection) from the beginning of the project ensuring that data gathering procedures conform to the Regulation (EU) 2016/679 of the European Parliament (GDPR) and of the Council of 27 April 2016. Each partner shall ensure compliance with any national legislation/regulation. A Security Advisory Board (SAB) will be responsible for ensuring that all measures are followed and fulfilled (primary focus on WP5) preventing data ending up in the hands of external unauthorised entities.

The EDM will assess the sensitivity of all deliverables before publication and will review progress regularly to assure that all deliverables are classified to the appropriate level. Furthermore, the EDM will provide recommendations for the pilot specifications based on defined legal/ethical/privacy aspects and will assess documentation from the pilot results for compliance. The EDM will outline high-level policy to be adapted to national conditions by each individual partner. The EDM will manage national adaptations by each partner, but following them will be the responsibility of each partner.

A Security Advisory Board (SAB) will be set up ensure that the measures defined under WP5 are followed and fulfilled preventing data ending up in the hands of external unauthorised entities (people, companies, etc.). The SAB will monitor activities or results that may cause security issues, and will manage those issues throughout the life of the project. The SAB will be composed of all WP5 task leaders, supported by the EDM. The board will have no executive powers but will perform as an advisory board for the Core Group members.

Main member organisations are ICOM, WEP, ATOS, SUST (EDM), however, all partners have resources allocated for participating in Task 1.2.

6 Data security

As mentioned under section 5, the EDM will collaborate with the **Security Advisory Board (SAB)** to ensure Data Security throughout the life of the project and for a period thereafter. The SAB will monitor activities or results that may cause security issues; it will manage those issues and ensure that provisions are in place for data security throughout the life of the project. This will include both the projects internal file repository and pilot data. Redundant data storage will be included to secure data recovery, but as design regarding pilot data is yet to be defined, this will be reviewed for the coming DMP.

6.1 GDPR

BD4OPEM will ensure the compliance of all project activities with good practices, as well as legal aspects of ethical, privacy and data protection issues from the beginning of the project, constantly monitoring the ethics and data management issues. This will ensure that data gathering procedures are based on consent forms and performed according to the GDPR Regulation (EU) 2016/679 and any relevant national legislation; “The principles of, and rules on the protection of natural persons with regard to the processing of their personal data should, whatever their nationality or residence, respect their fundamental rights and freedom, in particular their right to the protection of personal data” [2]. Legal, Ethical and Privacy (LEP) aspects will be integrated into the requirements elicitation. More complex data collection, such as consumer energy meter data will be considered as data collection sources become clearer.

The EDM will assure that local partners produce national adaptations and that each partner embraces the responsibility to follow them. Each partner needs to demonstrate that they follow national legislation.

BD4OPEM will be an easily replicable platform to be configured according to the characteristics and constraints of each country considering aspects like regulations and GDPR and the market specificities of each country.

7 Ethical aspects

The ethical aspects related to the personal data collected in this dataset are addressed in Task 1.4 (Ethics Management), where the processes of Protection of Personal Data (POPD) in BD4OPEM will be described.

Although Turkey does not have a ratified formal data protection law, BD4OPEM will comply with the European Data Protection Directive and adopt the higher standards of the EU. Most of the personal data needed for the project will be subject to local “Personal Data Protection Law” which is also known as KVKK (Kisisel Verilerin Korunmasi Kanunu) as an acronym in Turkish. KVKK is the equivalent form of GDPR which is applicable in European Union countries. Please refer to; <https://www.kvkk.gov.tr/Icerik/6649/Personal-Data-Protection-Law> for the applicable law. OEDAS is also responsible for a secondary regulation for data protection which takes part in Turkish Constitutional Law; https://global.tbmm.gov.tr/docs/constitution_en.pdf on page 15 in part "Privacy and protection of private life". OEDAS is responsible to EMRA (Energy Market Regulatory Authority) in Turkey. EMRA also releases its own directives to regulate the DSOs when needed. One of the directives restricts the share of data out of the country. However, R&D projects are mostly an exemption to those kinds of directives. OEDAS will decide which data is available and subject to communion.

No ethical issues are envisaged, but this will be monitored regularly during the project life cycle and as the data set develops. When collection and analysis of data from human activities are combined with other information for usage beyond the individual’s choice, it is at this intersection of the possible and desirable, that the moral and ethical use of big data become unclear. “The choices we make in acting upon that data have very real and very human consequences” [3].

8 Other issues

We will adhere to national legislations and adapt accordingly in discussions with each individual partner. It is the responsibility of each partner to identify and follow national legislation taking action where and when required.

BD4OPEM will be run in both Europe and Turkey. Our entire storage is in the EU region; therefore, we will be keeping all our data in countries with “adequate protection”. Users within the EU will be served by servers located in the EU, and the data collected from these users will also be stored in EU based servers. Moreover, EU based subjects will not have any means to access any other servers outside their continent, hence providing safe means to capsule EU data in the continent.

9 Conclusions

Good data management is a tool for supporting innovation and knowledge discovery. Equally important, it is a tool for knowledge integration and re-use.

The project recognises the importance of regulating data management issues and this document will evolve and be updated as the BD4OPEM project continues.

10 References

- [1] European Commission, "Data Management", https://ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/open-access-data-management/data-management_en.htm#A1-template.
- [2] E. Commission, Regulation (EU) 2016/679 of the European Parliament (GDPR) and of the Council of 27 April 2016
- [3] G. Booch, "*The Human and Ethical aspects of Big Data*", <https://ieeexplore.ieee.org/abstract/document/6750430>.



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 872525