The BRIGHT project is co-founded by the EU's Horizon 2020 innovation programme under grant agreement No 957816

Boosting DR through increased communIty-level consumer engaGement by combining Data-driven and blockcHain technology Tools with social science approaches and multi-value service design

Deliverable D1.2 Data Management Plan – first version

Author(s): Giuseppe Raveduto (ENG), Vincenzo Croce (ENG), Elena Sartini (CEL)

The project Boosting DR through increased communIty-level consumer engaGement by combining Data-driven and blockcHain technology Tools with social science approaches and multi-value service design (BRIGHT) has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957816. The sole responsibility for the content of this publication lies with the authors. It does not necessarily reflect the opinion of the Innovation and Networks Executive Agency (INEA) or the European Commission (EC). INEA or the EC are not responsible for any use that may be made of the information contained therein.

Imprint

Title:	**Data Management Plan – first version**
Contractual Date of Delivery to the EC:	30.04.2021
Actual Date of Delivery to the EC:	30.04.2021
Author(s):	Giuseppe Raveduto (ENG), Vincenzo Croce (ENG), Elena Sartini (CEL)
Participant(s):	Engineering, CyberEthics Lab.
Project:	Boosting DR through increased communIty-level consumer engaGement by combining Data-driven and blockcHain technology Tools with social science approaches and multi-value service design (BRIGHT)
Work Package:	WP1 – Project Management
Task:	T1.3 – Data Management Plan development
Confidentiality:	Public
Version:	1.0.0

Table of Contents

List of Tables

List of Acronyms and Abbreviations

BAG	Basisregistratie Addressen en Gebouwen (Key register of addresses and buildings)
BIM	Building Information Modeling
BMS	Building Management System
BRIGHT	Boosting DR through increased communIty-level consumer engaGement by combining Data-driven and blockcHain technology Tools with social science approaches and multi-value service design
CC	Creative Commons
CSV	Comma-Separated Values
DHW	Domestic Hot Water
DR	Demand-Response
EV	EV Electric Vehicle
FAIR	Findable, Accessible, Interoperable, and Reusable
HVAC	Heating, Ventilation and Air Conditioning
IoT	Internet of Things
MQTT	Message Queuing Telemetry Transport
PV	Photovoltaics
SAREF	Smart Applications REFerence
WP	Work Package

Table 1 List of Acronyms and Abbreviations

Executive Summary

The purpose of this document, *Data Management Plan – first version*, is to deliver the first version of the BRIGHT Data Management Plan describing the policies for the data to be collected, processed and/or generated during the course of the project.
The approach taken for this document referred to the *Guidelines on FAIR Data Management in Horizon 2020*: scientific research relies on advancements and improvements in earlier works made public, and FAIR data principles make it possible the scientific dialogue by ensuring that data are Findable, Accessible, Interoperable, and Reusable. The guidelines was used as a reference to define a dataset identification template and a list of questions that were circulated among the project partners to collect the necessary information.
This deliverable will be updated periodically to include the latest updates during the project life-cycle and, for this reason, it should be considered as a living document.

1 Introduction

Deliverable D1.2 is the first version of the Data Management Plan (DMP) for the BRIGHT project. This should be considered a living document, which means that updated versions will be created following the progress of the project, with two official releases already planned in M18 and M36. The current version reflects the status of the information currently available in the project.

1.1. Purpose

A DMP describes how the different datasets are stored and shared with third parties in order to fulfil the project goals. This document follows the guidelines specified in "*H2020 Programme Guidelines on FAIR Data Management in Horizon 2020*": the guidelines were used as a reference for the creation of a questionnaire sent to all the partners of the project to identify the datasets and the measures identified to make the data Findable, Accessible, Interoperable, and Reusable (FAIR).

1.2. Relation to Other Activities

Task 1.3, *Data Management Plan development*, defines the guidelines that will be implemented across the different Work Packages (WP) according to their specific objectives.
In addition, given its peculiarity, close collaboration with Task 1.4 is envisaged, especially with regard to privacy and data protection aspects.

1.3. Structure of the Document

The document is structured as follows: Section 1 – Introduction presents the document objectives and its structure. Section 2 – Dataset Collection includes the list of datasets generated within the project, categorizing them into:

- research data, produced or consumed by the project pilot sites to develop and validate the project tolls and technologies
- operational and observational data, collected to facilitate and monitor the project activities including the research ones.

Each dataset is described by a dedicated table and a list of related questions.
Section 3 – Findable, Accessible, Interoperable, and Reusable (FAIR) Data describes how each of the FAIR principles applies to the project, based on the analysis of the data collected from the project partners. Section 4 – Personal data protection's principles and Management of Personal Data describes the principles followed to ensure that personal data are processed following specific laws and regulations. Finally, Section 5 concludes the document and identifies possible next steps.

2 Dataset Collection

This section refers to the different categories of data acquired during the course of the project. The datasets are classified in two macro categories:

1. *Research Data,* collected specifically to develop or validate the project tools and technologies
2. *Operational and Observational Data,* collected to facilitate, enable, or monitor other project activities

Each of the dataset is defined by a table describing:

1. Dataset specific information
    a. Name
    b. Description
    c. Security & Privacy considerations

2. Information for each datatype in the dataset
    a. Datatype name
    b. Description
    c. Purpose
    d. Format
    e. Expected Size
    f. Source
    g. Access
    h. Recipients
    i. Metadata

In addition, for each dataset, the following additional questions were included:

1. How long will the data be stored/available?
2. Which data will be made openly available? If some data is kept closed provide rationale for doing so.
3. What methods or software tools are needed to access the data? Is documentation about the software needed to access the data included? Is it possible to include the relevant software (e.g. in open source code)? Is there any documentation available on the dataset?
4. Assess the interoperability of your data. Specify what data and metadata vocabularies, standards, or methodologies you will follow to facilitate interoperability.
5. Specify how the dataset will be licensed to permit the widest reuse possible. What types of license do you plan to use?
6. Specify when the data will be made available for re-use. If applicable, specify why and for what period a data embargo is needed.
7. Specify whether the data produced and/or used in the project is useable by third parties, in particular after the end of the project. If the re-use of some data is restricted, explain why.

The answers from each of the project partners and the dataset identification tables are reported below as received.

2.1 Research Data

Research data, grouped by country of origin

2.1.1 Belgium

2.1.1.1 Charging stations, Smart meters, and district heating data (Belgian Pilot)

Dataset Name	Charging stations
Description	Data collected from the pilot charging stations to understand their use (time of use, duration…)
Security & Privacy considerations	No personal data is included in this dataset
Contact Person	Chaim.de.mulder@openmotics.com
Datatype Name	- X.LP.Pe.totalpower - X.LP.Pe.powerphase1 - X.LP.Pe.powerphase2 - X.LP.Pe.powerphase3 - X.LP.I.currentphase1 - X.LP.I.currentphase2 - X.LP.I.currentphase3 - X.LP.U.voltagephase1 - X.LP.U.voltagephase2 - X.LP.U.voltagephase3 - *X.LP.-.numberofsessions* - *X.LP.dt.sessionxduration* - *X.LP.Pe.sessionxpower* - *X.LP.I.sessionxcurrent* - *X.LP.U.sessionxvoltage* *(Italics indicate time series that are not available yet and have not been assigned a final name)*
Description	- X.LP.Pe.totalpower: total power used by the system, to be subtracted from the maximum power that the charging stations are allowed to use. - X.LP.Pe.powerphase1/2/3: power on phase 1/2/3 - X.LP.I.currentphase1/2/3: current on phase 1/2/3 - X.LP.U.voltagephase1/2/3: voltage on phase 1/2/3 - *X.LP.-.numberofsessions: number of currently ongoing charging sessions* - *X.LP.dt.sessionxduration: duration up till now of session x* - *X.LP.Pe.sessionxpower: power delivered to session x* - *X.LP.I.sessionxcurrent: current delivered to session x* - *X.LP.U.sessionxvoltage: voltage delivered to session x*
Purpose	Build demand response control algorithms for charging station operation.
Format	Csv/zip files
Expected Size	Est. 50MB/month of data
Origin / Source	Charging station measurements (Powerdale device)
Access	Project partners will gain access to the data in case they require it for reaching their objectives as defined in the BRIGHT project proposal.

Recipients	Researchers
Metadata	Metadata will be created by providing, for each measured time series: - The name used for the series - Clarification on measurement location, interpretation, system context… - Measurement unit - Remarks specifying any anomalies in the data noticed by the data provider (in this case the responsible within DuCoop) This metadata will be gathered in a sheet and provided to the receiving partner along with the data itself.
Dataset Name	Smart meter data
Description	Data collected from the pilot smart meters (heat and electricity)
Security & Privacy considerations	Data will be anonimised, removing the link between a user's personal information and his/her smart meter readings, before sharing with any third party, within or outside of the Bright consortium
Contact Person	Chaim.de.mulder@openmotics.com
Datatype Name	xxxxxxxxx.energy App xx.x (to be anonimised)
Description	Data describing the electricity (App xx.x) and heat demand (xxxxxxxxx.energy) for the separate pilot households
Purpose	Assess possibilities for demand response (e.g. where are the largest gains possible by shifting demand)
Format	Csv/zip files
Expected Size	Est. 0.5 MB/user/month of data (current amount of users approx. 60)
Origin / Source	Smart meter measurements
Access	Project partners will gain access to the data in case they require it for reaching their objectives as defined in the BRIGHT project proposal.
Recipients	Researchers
Metadata	Metadata will be created by providing, for each measured time series: - The name used for the series - Clarification on measurement location, interpretation, system context… - Measurement unit - Remarks specifying any anomalies in the data noticed by the data provider (in this case the responsible within DuCoop) This metadata will be gathered in a sheet and provided to the receiving partner along with the data itself.
Dataset Name	District heating data
Description	Data collected from the district heating network and related assets (industrial waste heat heat exchanger, heat pump, biogas boiler)
Security & Privacy considerations	Data relating to industrial processes can be sensitive and should be handled with care when sharing with partners and/or the public.
Contact Person	Chaim.de.mulder@openmotics.com
Datatype Name	*For general district heating network* IND1.WNET.T.Buffervat Boven/T_Buffervat_boven IND1.WNET.T.Buffervat Midden/T_Buffervat_midden IND1.WNET.T.Buffervat Onder/T_Buffervat_onder

	IND1.WNET.T.Aankomst/T_Vetrek_Warmtenet IND1.WNET.T.Vertrek/T_Retour_Warmtenet IND1.WNET.Q.Debiet warmtenet retour/Debiet_van_CV IND1.WNET.Ew.kWh Warmtewisselaar Restwarmte Christeyns MSB IND1.WNET.Ew.kWh Warmtewisselaar Ketels Christeyns MSB FAA.WNET.T.Buffervat 1/T_Buffervat_1 FAA.WNET.T.Buffervat 2/T_Buffervat_2 FAA.WNET.T.Buffervat 3/T_Buffervat_3 FAA.WNET.T.Vertrek Gebouwen/T_Vertrek_Gebouwen FAA.WNET.dT.Cal Gebouwen/T_Diff_Cal_Gebouwen FAA.WNET.T.Cal Gebouwen In/T_Inlet_Cal_Gebouwen FAA.WNET.Ew.kWh Calorimeter Gebouwen MSB FAA.WNET.dT.Cal totaal/T_Diff_Cal_Christeyns FAA.WNET.T.Cal totaal In/T_Inlet_Cal_Christeyns FAA.WNET.Ew.kWh Calorimeter Totaal MSB FAA.WNET.dT.Cal WZI/T_Diff_Cal_Waterzuivering FAA.WNET.T.Cal WZI In/T_Inlet_Cal_Waterzuivering FAA.WNET.Ew.kWh Calorimeter Waterzuivering MSB FAA.WNET.dT.Cal WPO/T_Diff_Cal_Warmtepomp FAA.WNET.T.Cal WPO In/T_Inlet_Cal_Warmtepomp FAA.WNET.Ew.kWh Calorimeter Warmtepomp MSB FAA.WNET.dT.Cal kantoren/T_Diff_Cal_Kantoren FAA.WNET.T.Cal kantoren In/T_Inlet_Cal_Kantoren FAA.WNET.Ew.kWh Calorimeter Kantoren MSB FAA.WNET.dT.Cal Biogasboiler/T_Diff_Cal_Biogasboiler FAA.WNET.T.Cal Biogasboiler In/T_Inlet_Cal_Biogasboiler FAA.WNET.Ew.kWh Calorimeter Biogasboiler MSB FAA.WNET.Ew.kWh Caloriemeter School MSB FAA.WNET.Ew.kWh Caloriemeter Sanitair sporthal MSB FAA.WNET.Ew.kWh Caloriemeter Ventilatie sporthal MSB *For individual living units* district_heating_temperature district_heating_return_temperature district_heating_flow Thermostat_hsf_station ch_supply_limit Thermostat_living_unit Temperature_living_unit Measured temperature Predicted temperature
Description	Temperature, flow and energy for heat sources and heat sinks; Temperature in different buffer tanks; temperature, flows in individual heat exchangers;

	thermostat state and room temperature in individual living units; outside temperature
Purpose	Assess possibilities for optimised operation of the district heating network, including user-side demand response
Format	Csv/zip files
Expected Size	Est. 8-10 MB/month of district heating data, plus est. 4 MB/user/month of individual data
Origin / Source	Measurements in place for monitoring and control of the district heating network; measurements inside the individual heat exchangers (HSF); measurements inside living units; weather station on rooftop of a pilot building.
Access	Project partners will gain access to the data in case they require it for reaching their objectives as defined in the BRIGHT project proposal.
Recipients	Researchers
Metadata	Metadata will be created by providing, for each measured time series: - The name used for the series - Clarification on measurement location, interpretation, system context… - Measurement unit - Remarks specifying any anomalies in the data noticed by the data provider (in this case the responsible within DuCoop) This metadata will be gathered in a sheet and provided to the receiving partner along with the data itself.

Table 2 Charging stations, smart meters, and district heating data (Belgian pilot)

1) How long will the data be stored/available?

The data will be stored by DuCoop for as long as required to provide its services to the pilot residents, and at least for the duration of the project.

2) Which data will be made openly available? If some data is kept closed provide rationale for doing so.

No personal data will be made openly available outside of the BRIGHT consortium.
Non-personal data will be made available to the public in the form of published datasets where possible.

3) What methods or software tools are needed to access the data? Is documentation about the software needed to access the data included? Is it possible to include the relevant software (e.g. in open source code)? Is there any documentation available on the dataset?

The data is saved in InfluxDB databases and can (by pilot partners) be accessed by using Grafana, a browser-based, open-source visualisation tool, or by using an API provided by OpenMotics, the technology provider at the pilot site.
In the case of publicly available data, it will be possible for a user to download a dataset and use his/her preferred data analsyis tools.
All information regarding the data (naming convention, location of sensors, measurement units…) is available, but not aggregated in an easily accessible format. The required information on the datasets will be provided ad-hoc to project partners that require the data. Where possible, metadata will be made publicly available along with the actual datasets.

4) **Assess the interoperability of your data. Specify what data and metadata vocabularies, standards, or methodologies you will follow to facilitate interoperability.**

De Nieuwe Dokken pilot (both operator DuCoop and technology provider OpenMotics) are involved in the EU H2020 InterConnect project, aimed specifically at interoperability using SAREF protocols. All interoperability-related methodologies will orginate from that project.

5) **Specify how the dataset will be licensed to permit the widest reuse possible. What types of license do you plan to use?**

Any datasets made publicly available will be licensed under a CC BY-NC 4.0 license (https://creativecommons.org/licenses/by-nc/4.0/)

6) **Specify when the data will be made available for re-use. If applicable, specify why and for what period a data embargo is needed.**

Data will be made publicly available on a half-yearly to yearly basis, mainly based on efficiency considerations (i.e. monthly would lead to a large amount of published datasets, longer than yearly would lead to very large datasets and long preparation times).

7) **Specify whether the data produced and/or used in the project is useable by third parties, in particular after the end of the project. If the re-use of some data is restricted, explain why.**

All publicly published data is of course useable by third parties. Re-use is restricted in the ways specified in the used license (see above).

2.1.1.2 Residential Smart Meter Data (Belgian Pilot)

Dataset Name	*Residential smart meter data*
Description	*Smart meter data offered by the Flemish DSO Fluvius*
Security & Privacy considerations	*Anonomized, data cannot be shared with other partners without signing an NDA with the data owner (Fluvius)*
Contact Person	*Matthias.strobbe@ugent.be*
Datatype Name	*Residential consumption profiles*
Description	*Net consumption and injection measurements for houses with a 15 minute resolution.*
Purpose	*Data was measured during the first proof-of-concept phase of the roll-out of smart meters in Flanders, in the period 2010-2014.*
Format	*CSV files*
Expected Size	*±38 GB*
Origin / Source	*Backend system Fluvius*
Access	*The data is accessible for selected researchers after signing an NDA with Fluvius.*
Recipients	*Researchers*
Metadata	*A detailed description of the data is available in a Word document.*

Table 3 Residential smart meter data (Belgian pilot)

1) **How long will the data be stored/available?**

There is no end data set for using the dataset for research purposes.

2) **Which data will be made openly available? If some data is kept closed provide rationale for doing so.**

The data can only be used after signing an NDA with the data owner Fluvius, so the data will not be made publicly available.

3) **What methods or software tools are needed to access the data? Is documentation about the software needed to access the data included? Is it possible to include the relevant software (e.g. in open source code)? Is there any documentation available on the dataset?**

The data consist of csv files which can be easily processed by all kinds of tools.

4) **Assess the interoperability of your data. Specify what data and metadata vocabularies, standards, or methodologies you will follow to facilitate interoperability.**

The data consists of csv files with a limited amount of columns and is easy to understand and process.

5) **Specify how the dataset will be licensed to permit the widest reuse possible. What types of license do you plan to use?**

We cannot license the data as we are not the owner of the data.

6) **Specify when the data will be made available for re-use. If applicable, specify why and for what period a data embargo is needed.**

The data can be reused for other research goals, but only after an agreement with Fluvius and signing an NDA.

7) **Specify whether the data produced and/or used in the project is useable by third parties, in particular after the end of the project. If the re-use of some data is restricted, explain why.**

The data can be reused by other partiers and for other research goals, but only after an agreement with Fluvius and signing an NDA.

2.1.1.3 Residential Flexibility Data

Dataset Name	*Residential flexibility data*
Description	*Consumption profiles and flexibility information from residential households devices (mainly whitegoods), collected during the Flemish research project Linear (https://www.energyville.be/onderzoek/linear-actieve-vraagsturing-bij-gezinnen).*
Security & Privacy considerations	*The data is anonimised.*
Contact Person	*Matthias.strobbe@ugent.be*
Datatype Name	*Residential flexibility data.*
Description	*Consumption profiles from households and some selected appliances in the homes, PV production profiles for about half of the households and flexibility information from smart appliances (mainly whitegoods), i.e., when the appliances were programmed, deadline by when they should be ready and actual moment that they were activated.*
Purpose	*The data was collected as part of a large research and*

	demonstration project in Flanders (Linear) on the technical and economical feasibility of residential demand response services.
Format	*Data is available in a relational database and can be exported as csv files.*
Expected Size	*±5GB as compressed database dump file*
Origin / Source	*MySQL database that was used as backend for the research project.*
Access	*The data can only be used by partners directly involved in the Linear project.*
Recipients	*Researchers*
Metadata	*Metadata can be provided in a separate document.*

Table 4 Residential flexibility data

1) **How long will the data be stored/available?**

The data is available without specific end date for researchers involved in the Linear project.

2) **Which data will be made openly available? If some data is kept closed provide rationale for doing so.**

It's not allowed to make any of the data openly available. However we built some statistical models on top of a subset of the data. These models were published in scientific publications (e.g. http://users.atlantis.ugent.be/cdvelder/papers/2016/sadeghianpourhamami2016apen.pdf) and could be used as a starting point to create a synthetic data generator.

3) **What methods or software tools are needed to access the data? Is documentation about the software needed to access the data included? Is it possible to include the relevant software (e.g. in open source code)? Is there any documentation available on the dataset?**

The data is available in a relational database (MySQL) and can be easily exported as CSV files.

4) **Assess the interoperability of your data. Specify what data and metadata vocabularies, standards, or methodologies you will follow to facilitate interoperability.**

No specific standards or metadata vocabularies were used, but the structure of the database is quite self-explanatory and extra explanation could be provided if needed.

5) **Specify how the dataset will be licensed to permit the widest reuse possible. What types of license do you plan to use?**

It's not allowed to license the data to partners not involved in the Linear project.

6) **Specify when the data will be made available for re-use. If applicable, specify why and for what period a data embargo is needed.**

The data can only be reused by partners involved in the Linear project. However, based on the published statistical models representing the available flexibility in the pilot, synthetic data generators could be developed for diverse applications.

7) **Specify whether the data produced and/or used in the project is useable by third parties, in particular after the end of the project. If the re-use of some data is restricted, explain why.**

The data can only be reused by partners involved in the Linear project. However, based on the published statistical models representing the available flexibility in the pilot, synthetic data generators could be developed for diverse applications.

2.1.1.4 Building BMS & Sensor Data

Dataset Name	*Building BMS & sensor data*
Description	*BMS and IoT sensor data collected from an 12-storey office building and residential living lab.*
Security & Privacy considerations	*The data is anonimised.*
Contact Person	*Matthias.strobbe@ugent.be*
Datatype Name	*BMS, sensor & BIM data office building*
Description	- *BMS data: measurements for many aspects of the HVAC system, weather info, context data within the building (room temperatures, status of remotely controllable blinds and windows, some air quality measurements) since June 2016 (and ongoing)* - *IoT sensor data: extra measurements on air quality, noise, radiator valve status, status of windows and doors in the in the office & meeting rooms and in the corridors for floors 9-12 since November 2020 (some sensors still need to be deployed)* - *BIM model of the whole building*
Purpose	*BMS data is used for operational management of the building. BMS + sensor data is also used for an internal research project.*
Format	*Historical data is available in a relational database and could be exported as CSV files. Recent and real-time data is available in an Influx based time series database and be exported as CSV files or requested as JSON data via an API.*
Expected Size	*5MB/day as zipped CSV file*
Origin / Source	*BMS system and research data platform (for the IoT sensor data)*
Access	*Data can be provided to interested parties inside and outside the consortium for research purposes (after internal approval of the request).*
Recipients	*Researchers*
Metadata	*Metadata is available in JSON format.*
Datatype Name	*Sensor, actuator and BIM data for residential living lab.*
Description	*Historical and real-time data is available for about 2.5 years for many environmental parameters (temperature, air quality, user presence, light intensity, weather info, detailed energy consumptions, …) and status info on many building systems (HVAC, blinds, windows, doors, curtains, lights, …).*
Purpose	*Data is collected to support the research activities in the living lab.*
Format	*Historical and real-time data is available in an Influx based time series database and be exported as CSV files or requested as JSON data via an API.*
Expected Size	*14MB/day as zipped CSV file*
Origin / Source	*The data from the living lab is collected in a research data platform.*
Access	*Data can be provided to interested parties inside and outside the consortium for research purposes (after internal approval of the request).*

Recipients	*Researchers*
Metadata	*Metadata is available in JSON format.*

Table 5 Building BMS & sensor data

1) **How long will the data be stored/available?**

The data is available without specific end date for research purposes.

2) **Which data will be made openly available? If some data is kept closed provide rationale for doing so.**

There are currently no concrete plans to make data available to the general public, but data can be shared for research purposes.

3) **What methods or software tools are needed to access the data? Is documentation about the software needed to access the data included? Is it possible to include the relevant software (e.g. in open source code)? Is there any documentation available on the dataset?**

Data can be shared as CSV files or in JSON format and is thus easy to process.

4) **Assess the interoperability of your data. Specify what data and metadata vocabularies, standards, or methodologies you will follow to facilitate interoperability.**

No specific standards or metadata vocabularies were used, but the structure of the data is quite self-explanatory and extra explanation can be provided if needed.

5) **Specify how the dataset will be licensed to permit the widest reuse possible. What types of license do you plan to use?**

Data can be shared with interested parties for research purposes after approval of the request and for the agreed purposes. Possible in the future some subsets of the data will be openly released to the research community.

6) **Specify when the data will be made available for re-use. If applicable, specify why and for what period a data embargo is needed.**

Data can be shared and reused for other research applications on request.

7) **Specify whether the data produced and/or used in the project is useable by third parties, in particular after the end of the project. If the re-use of some data is restricted, explain why.**

Data can be shared and reused for other research applications on request, both for partners inside and outside the consortium.

2.1.2 Greece

2.1.2.1 Electricity Submeter, Indoor Conditions, Home Monitoring Data

Dataset Name	*Electricity submeter data*
Description	*Data collected from pilot electricity submeters will be used to construct energy profile and visualization tools*
Security & Privacy considerations	*Data will be anonymised, removing the link between a user's personal information. Moreover, will comply with GDPR specifications*
Contact Person	dpo@watt-volt.gr
Datatype Name	*submeter_data*

Description	*Data analysis for describing electricity measures with the most understoodable way*
Purpose	*Visualization of historical and real time data in order to give consumers the ability to monitor and understand their electric consumption empowering them to achieve energy savings through DR recommendations*
Format	*To be specified during pilot development*
Expected Size	*TBD*
Origin / Source	*Electricity submeter infrastructure*
Access	*Only project partners will gain access to the data in case they require it for reaching their objectives as defined in the BRIGHT project proposal.*
Recipients	*TBD*
Metadata	Metadata will be created for each time serie measurement: • The name used for the series • Measurement location, timestamp • Measurement unit
Dataset Name	*Indoor conditions data (temperature, humidity)*
Description	*Data collected from the IoT devices will be used for real time monitoring and visualization of historical data*
Security & Privacy considerations	*Data will be anonymised, removing the link between a user's personal information. Moreover, will comply with GDPR specifications*
Contact Person	dpo@watt-volt.gr
Datatype Name	*temperature_data, humidity_data*
Description	*Visualization of indoor parameters (temperature, humidity)*
Purpose	*Give customers the ability to monitor their household conditions*
Format	*To be specified during pilot development*
Expected Size	*TBD*
Origin / Source	*IoT devices*
Access	*Only project partners will gain access to the data in case they require it for reaching their objectives as defined in the BRIGHT project proposal.*
Recipients	*TBD*
Metadata	Metadata will be created for each time series measurement: • The name used for the series • Measurement location, timestamp • Measurement unit
Dataset Name	Home usage patterns (door contacts, human presence)
Description	*Data collected from the IoT devices will be used for real time monitoring*
Security & Privacy considerations	*Data will be anonymised, removing the link between a user's personal information. Moreover, will comply with GDPR specifications*
Contact Person	dpo@watt-volt.gr
Datatype Name	window_status, human_presence
Description	*Real time monitoring of boolean values*

Purpose	*Give customers ability to correlate indoor conditions with window status and human presence with security*
Format	*To be specified during pilot development*
Expected Size	TBD
Origin / Source	*IoT devices*
Access	*Only project partners will gain access to the data in case they require it for reaching their objectives as defined in the BRIGHT project proposal.*
Recipients	TBD
Metadata	Metadata will be created for each time series measurement: • The name used for the series • Measurement location, timestamp
Dataset Name	Home automation and monitoring of energy consumption in appliance level (smart plug, smart relay)
Description	Data collected from pilot IoT devices *will be used for real time monitoring and visualization tools*
Security & Privacy considerations	*Data will be anonymised, removing the link between a user's personal information. Moreover, will comply with GDPR specifications*
Contact Person	dpo@watt-volt.gr
Datatype Name	*smaprt_plug, smart_relay*
Description	*Data analysis for describing electricity measures with the most understoodable way and provide remote control*
Purpose	*Visualization of historical and real time data in order to give consumers the ability to monitor and understand their electric consumption empowering them to remote control of heavy consuming appliances and achieve energy savings through DR recommendations*
Format	*To be specified during pilot development*
Expected Size	TBD
Origin / Source	*IoT devices*
Access	*Only project partners will gain access to the data in case they require it for reaching their objectives as defined in the BRIGHT project proposal.*
Recipients	TBD
Metadata	Metadata will be created for each time series measurement: • The name used for the series • Measurement location, timestamp • Measurement unit

Table 6 Electricity submeter, indoor conditions, home monitoring data

1) How long will the data be stored/available?

The data will be stored as long as required to provide its services to the pilot residents, and at least for the duration of the project.

2) Which data will be made openly available? If some data is kept closed provide rationale for doing so.

No data will be publicly available.

3) **What methods or software tools are needed to access the data? Is documentation about the software needed to access the data included? Is it possible to include the relevant software (e.g. in open source code)? Is there any documentation available on the dataset?**

Not available yet.

4) **Assess the interoperability of your data. Specify what data and metadata vocabularies, standards, or methodologies you will follow to facilitate interoperability.**

Not available yet.

5) **Specify how the dataset will be licensed to permit the widest reuse possible. What types of license do you plan to use?**

Not available yet.

6) **Specify when the data will be made available for re-use. If applicable, specify why and for what period a data embargo is needed.**

Not available yet.

7) **Specify whether the data produced and/or used in the project is useable by third parties, in particular after the end of the project. If the re-use of some data is restricted, explain why.**

Not available yet.

2.1.2.2 Residential Space Heating and DHW

Dataset Name	*Residential space heating and DHW preparation data for gas boilers*
Description	*Data collected rom gas boiler, thermostats and smartphone user applications related with space heating and DHW preparation scenarios of residential consumers.*
Security & Privacy considerations	*No personal data are included in this dataset*
Contact Person	*Stratos Keranidis (stratos@mydomx.eu)*
Datatype Name	• *blr_mod_lvl* • *blr_t* • *dhw_t* • *flame* • *heat* • *water* • *t_out* • *t_r* • *rh_r* • *t_set* • *t_r_set* • *t_dhw_set* • *heat_set* • *water_set* • *otc_cur* • *bypass*

<table>
<tr><td></td><td>

- *pid*

</td></tr>
<tr><td>Description</td><td>
<table>
<tr><td>blr_mod_lvl</td><td>Current Boiler modulation Level (as percentage of max boiler output, with most common value being 24 kW)</td></tr>
<tr><td>blr_t</td><td>Current Boiler Water temperature</td></tr>
<tr><td>dhw_t</td><td>Current Domestic Hot Water Temperature</td></tr>
<tr><td>flame</td><td>Current Boiler Flame State - Shows whether the boiler is ignited</td></tr>
<tr><td>heat</td><td>Current Boiler Heat State - Shows whether the boiler circulator is active</td></tr>
<tr><td>water</td><td>Current Boiler Water State - Shows whether the boiler DHW is active</td></tr>
<tr><td>t_out</td><td>Outdoor temperature - Input taken from the domX GW temperature Sensor (default) or by the Boiler temperature probe if it exists</td></tr>
<tr><td>t_r</td><td>Current Room Temperature - Reported by the Thermostat or the domX indoor climate sensor</td></tr>
<tr><td>rh_r</td><td>Current Room Humidity - Reported by the domX indoor climate sensor</td></tr>
<tr><td>t_set</td><td>Desired Boiler Water temperature Setting - Set by the Thermostat or the DomX GW</td></tr>
<tr><td>t_r_set</td><td>Desired Room Temperature Setting - Set by the Thermostat or the DomX GW</td></tr>
<tr><td>t_dhw_set</td><td>Desired DHW Temperature Setting - Set by the Thermostat or the DomX GW</td></tr>
<tr><td>heat_set</td><td>Desired Boiler Heat Setting - Set by the Thermostat or the DomX GW (Enabled/Disabled)</td></tr>
<tr><td>water_set</td><td>Desired Boiler Water Setting - Set by the Thermostat or the DomX GW (Enabled/Disabled)</td></tr>
<tr><td>otc_cur</td><td>Weather compensation tradeoff that adapts the aggressiveness of the heating control algorithm (Controls the MAX boiler water temperature to be set)</td></tr>
<tr><td>bypass</td><td>Control the boiler temperature controller source (0: default, 1: weather compensation with user assigned otc_cur, 2: fixed boiler temperature, 4: weather compensation with cloud controlled otc_cur)</td></tr>
<tr><td>pid</td><td>Boiler temperature as calculated by the heating control algorithm (assigned to the t_set parameter if bypass equals 2 or 4)</td></tr>
</table>
</td></tr>
<tr><td>Purpose</td><td>Monitoring and control of residential space heating and DHW preparation gas boilers</td></tr>
<tr><td>Format</td><td>csv, json</td></tr>
<tr><td>Expected Size</td><td>2 MBs per home per day</td></tr>
<tr><td>Origin / Source</td><td>DomX heating controller and smartphone application</td></tr>
</table>

Access	*Project partners will gain access to the data in case they require it for reaching their objectives as defined in the BRIGHT project proposal.*
Recipients	*Researchers*
Metadata	*Additional metadata can be offered to specify the:* - *approximate location per home* - *boiler vendor and model per home* - *measurement units per data type* - *description per data type* - *data collection intervals per data type* *These metadata can be gathered in a csv file and provided to the receiving partner along with the data itself.*

Table 7 Residential space heating and DHW preparation data for gas boilers

1) **How long will the data be stored/available?**

These data are being continuously collected for delivering end user services to the respective pilot participants, covering at least the project duration and two years after.

2) **Which data will be made openly available? If some data is kept closed provide rationale for doing so.**

No personal data will be made openly available outside of the BRIGHT consortium.
Non-personal data will be made available to the public in the form of published datasets where possible.

3) **What methods or software tools are needed to access the data? Is documentation about the software needed to access the data included? Is it possible to include the relevant software (e.g. in open source code)? Is there any documentation available on the dataset?**

These data can be shared through exported csv/ json files. It is also possible to access real-time data through the domX REST API, on an ad-hoc basis and when required by a respective project partner.

4) **Assess the interoperability of your data. Specify what data and metadata vocabularies, standards, or methodologies you will follow to facilitate interoperability.**

For the moment, we do not follow any standard interoperability protocol, but we plan to support SAREF.

5) **Specify how the dataset will be licensed to permit the widest reuse possible. What types of license do you plan to use?**

This specific dataset is not reusable and will not be licensed for reuse.

6) **Specify when the data will be made available for re-use. If applicable, specify why and for what period a data embargo is needed.**

N/A

7) **Specify whether the data produced and/or used in the project is useable by third parties, in particular after the end of the project. If the re-use of some data is restricted, explain why.**

No, this kind of data cannot be used for other activities, especially from third parties not belonging to the Consortium.

2.1.3 Italy

2.1.3.1 Charging Station Data (Italian Pilot)

Dataset Name	*Charging Station Data*
Description	*Data collected from charging stations deployed in Terni pilot site*
Security & Privacy considerations	*Anonymisation*
Contact Person	*Francesco Bellesini (francesco.bellesini@emotion-team.com)*
Datatype Name	*Charging Station ID* *Charging Station Electric Current* *Socket ID* *Socket Status* *Charging Session ID* *Charging session Start Time* *Charging session End Time* *Charging session Energy* *Charging session Cost*
Description	*Historical and real-time data related to charging stations involved in Terni pilot site demonstration activities*
Purpose	*Validate intelligent centralized flexibility aggregation by EV fleet operator by smart recharging to better match with flexibility profile request*
Format	*JSON*
Expected Size	*100 KBs per device per day*
Origin / Source	*Data will be provided by EMOT e-Mobility platform*
Access	*Access will be provided to consortium partners according to project objectives*
Recipients	*Project partners; Public access (for anonymized data)*
Metadata	Metadata will be created based on OCPP

Table 8 Charging station data (Italian Pilot)

1) How long will the data be stored/available?

5 years

2) Which data will be made openly available? If some data is kept closed provide rationale for doing so.

Data will be anonymized and after they will be made available

3) What methods or software tools are needed to access the data? Is documentation about the software needed to access the data included? Is it possible to include the relevant software (e.g. in open source code)? Is there any documentation available on the dataset?

Data access methods are:

- API or MQTT broker for real time data;
- CSV file for historical data.

4) Assess the interoperability of your data. Specify what data and metadata vocabularies, standards, or methodologies you will follow to facilitate interoperability.

To facilitate interoperability, every data provided will be in JSON format

5) **Specify how the dataset will be licensed to permit the widest reuse possible. What types of license do you plan to use?**

Any datasets made publicly available will be licensed under a CC BY-NC 4.0 license (https://creativecommons.org/licenses/by-nc/4.0/)

6) **Specify when the data will be made available for re-use. If applicable, specify why and for what period a data embargo is needed.**

Data will be made publicly available on a yearly basis

7) **Specify whether the data produced and/or used in the project is useable by third parties, in particular after the end of the project. If the re-use of some data is restricted, explain why.**

Published data is useable by third parties according to CC BY-NC 4.0 license

2.1.3.2 Electric Vehicle Data (Italian Pilot)

Dataset Name	*Electric Vehicle Data*
Description	*Data collected from electric vehicles deployed in Terni pilot site*
Security & Privacy considerations	*Anonymisation*
Contact Person	*Francesco Bellesini (francesco.bellesini@emotion-team.com)*
Datatype Name	*Electric Vehicle ID* *Electric Vehicle Model* *Connector Type* *Battery Capacity* *Battery Power* *Timestamp* *SoC* *Latitude* *Longitude* *Speed* *Kilometers Autonomy* *Odometer*
Description	*Historical and real-time data related to electric vehicles involved in Terni pilot site demonstration activities*
Purpose	*Validate intelligent centralized flexibility aggregation by EV fleet operator by smart recharging to better match with flexibility profile request*
Format	*JSON*
Expected Size	*100 KBs per device per day*
Origin / Source	*Data will be provided by EMOT e-Mobility platform*
Access	*Access will be provided to consortium partners according to project objectives*
Recipients	*Project partners; Public access (for anonymized data)*
Metadata	Metadata will be created based on Python

Table 9 Electric Vehicle Data (Italian Pilot)

1) **How long will the data be stored/available?**

5 years

2) **Which data will be made openly available? If some data is kept closed provide rationale for doing so.**

Data will be anonymized and after they will be made available

3) **What methods or software tools are needed to access the data? Is documentation about the software needed to access the data included? Is it possible to include the relevant software (e.g. in open source code)? Is there any documentation available on the dataset?**

Data access methods are:

- API or MQTT broker for real time data;
- CSV file for historical data.

4) **Assess the interoperability of your data. Specify what data and metadata vocabularies, standards, or methodologies you will follow to facilitate interoperability.**

To facilitate interoperability, every data provided will be in JSON format

5) **Specify how the dataset will be licensed to permit the widest reuse possible. What types of license do you plan to use?**

Any datasets made publicly available will be licensed under a CC BY-NC 4.0 license (https://creativecommons.org/licenses/by-nc/4.0/)

6) **Specify when the data will be made available for re-use. If applicable, specify why and for what period a data embargo is needed.**

Data will be made publicly available on a yearly basis

7) **Specify whether the data produced and/or used in the project is useable by third parties, in particular after the end of the project. If the re-use of some data is restricted, explain why.**

Published data is useable by third parties according to CC BY-NC 4.0 license

2.1.3.3 Apartment building, ASM headquarters, users/prosumers data (Italian Pilot)

Dataset Name	*ASM dataset*
Description	*From 2018 to 2021*
Security & Privacy considerations	*Anonymisation*
Contact Person	*Tommaso Bragatto (tommaso.bragatto@asmterni.it)*
Datatype Name	***Apartment building level data***
Description	*Energy data regarding the IoT smart meters of a customer group of an apartment building.*
Purpose	*Data collected to improve energy performances of customers.*
Format	*CSV/JSON*
Expected Size	*1MB/day,user*
Origin / Source	*MQTT Broker*
Access	*Describe the access policy for both project partners and general public*
Recipients	*NoSQL db in ASM server farm*
Metadata	*N/A*

Datatype Name	***ASM headquarters data***
Description	*Data from energy units, like photovoltaic plant.*
Purpose	*Data collected to manage grid unbalances.*
Format	*CSV/JSON*
Expected Size	*1MB/day*
Origin / Source	*MQTT Broker*
Access	*Access to ASM R&D unit*
Recipients	*NoSQL db in ASM server farm*
Metadata	*N/A*
Datatype Name	***Users/Prosumers***
Description	*Data from customers, like energy, voltages, currents.*
Purpose	*Data collected to study customers consumption patters.*
Format	*CSV/JSON*
Expected Size	*1MB/day,user*
Origin / Source	*MQTT Broker*
Access	*Access to ASM R&D unit*
Recipients	*NoSQL db in ASM server farm*
Metadata	*N/A*

Table 10 ASM dataset

1) **How long will the data be stored/available?**

Data related to the three datasets are stored for 5 years.

2) **Which data will be made openly available? If some data is kept closed provide rationale for doing so.**

No data are openly available.

3) **What methods or software tools are needed to access the data? Is documentation about the software needed to access the data included? Is it possible to include the relevant software (e.g. in open source code)? Is there any documentation available on the dataset?**

Data are avilable through MQTT protocol. Documentation is available.

4) **Assess the interoperability of your data. Specify what data and metadata vocabularies, standards, or methodologies you will follow to facilitate interoperability.**

A partial list of standard used includes:

- OpenADR 2.0;
- ASCII (American Standard Code for Information Interchange);
- MQTT;
- RESTful services.

5) **Specify how the dataset will be licensed to permit the widest reuse possible. What types of license do you plan to use?**

Data will be stored in servers, either on site of the pilots or on locations indicated by the technology provider. Public data will be published after the release of the respective deliverable or after the end of the project.

6) **Specify when the data will be made available for re-use. If applicable, specify why and for what period a data embargo is needed.**

Data have been made available since project start.

7) **Specify whether the data produced and/or used in the project is useable by third parties, in particular after the end of the project. If the re-use of some data is restricted, explain why.**

The availability of data after the end of the project depends highly on the type and content of the data. Therefore, storing data on a public platform needs to be discussed with the contributor of the data.

2.1.4 Netherlands

2.1.4.1 Elaad EV Charging Data

Dataset Name	*ELaad EV charging data*
Description	*EV charging sessions collected from public charging infrastructure in The Netherlands.*
Security & Privacy considerations	*The data is anonimised.*
Contact Person	*Matthias.strobbe@ugent.be*
Datatype Name	*EV charging sessions*
Description	*Recorded connection times, charging times, power consumption and location info.*
Purpose	*Data collected to monitor, maintain and improve charging infrastructure.*
Format	*RDA (can be converted to CSV)*
Expected Size	*200MB*
Origin / Source	*EV charging infrastructure backend server*
Access	*The original data is only accessible for imec, not for project partners nor the general public. However a synthetic data generator based on this data was created which is freely available (https://github.com/mlahariya/EV-SDG). A subset of the dataset is also publicly available via https://platform.elaad.io/download-data/*
Recipients	*Researchers*
Metadata	*The data is quite self-explanatory, but extra metadata can be provided for the data generated by the synthetic data generator on request.*

Table 11 Elaad EV charging data

1) **How long will the data be stored/available?**

There is no end data set for using the dataset and synthetic data generator for research purposes.

2) **Which data will be made openly available? If some data is kept closed provide rationale for doing so.**

The original data is only accessible for imec, not for project partners nor the general public, as this is not allowed by the data owner (ELaad.nl). However a synthetic data generator based on this data was created which is freely available (https://github.com/mlahariya/EV-SDG). A subset of the dataset is also available to the general public via https://platform.elaad.io/download-data/ .

3) **What methods or software tools are needed to access the data? Is documentation about the software needed to access the data included? Is it possible to include the relevant software (e.g. in open source code)? Is there any documentation available on the dataset?**

The synthetic data generator generates data as csv files which can be easily processed by all kinds of tools.

4) **Assess the interoperability of your data. Specify what data and metadata vocabularies, standards, or methodologies you will follow to facilitate interoperability.**

The resulting data for the synthetic data generator are self-explanatory csv files.

5) **Specify how the dataset will be licensed to permit the widest reuse possible. What types of license do you plan to use?**

We cannot license the original data as we are not the owner of the data. The synthetic data generator is released as open source software (MIT license).

6) **Specify when the data will be made available for re-use. If applicable, specify why and for what period a data embargo is needed.**

The original data will not be made available for re-use, but the synthetic data generator is freely available for any kind of application.

7) **Specify whether the data produced and/or used in the project is useable by third parties, in particular after the end of the project. If the re-use of some data is restricted, explain why.**

The original data cannot be reused, but the synthetic data generator can be freely used by any interested third party and for any application.

2.1.4.2 ElaadNL Open Datasets for Electric Mobility Research

Dataset Name	*ElaadNL Open Datasets for Electric Mobility Research*
Description	ElaadNL offers a unique data and information package regarding electric vehicle (EV) charging patterns based on real world measurements.
Security & Privacy considerations	*Public domain data*
Contact Person	*Elaad, data@elaad.nl/ Nazir Refa* nazir.refa@elaad.nl
Datatype Name	*Elaadnl_open_ev_datasets*
Description	*Overview of 10k random charging events including 15 minutes metervalues per transaction*
Purpose	Based on these data sets one should be able to run, and validate simulations and conduct studies on (future) charging behaviour of EVs.
Format	*.csv/.xls*
Expected Size	*18 MB*
Origin / Source	*https://platform.elaad.io/analyses/ElaadNL_opendata.php*
Access	*Open*
Recipients	*Researchers*
Metadata	*None*

Table 12 ElaadNL open datasets for electric mobility research

1) **How long will the data be stored/available?**

This dataset is available till 2023-12-31, here.

2) Which data will be made openly available? If some data is kept closed provide rationale for doing so.

In addition to Overview of 10k random charging events including 15 minutes metervalues per transaction, distribution of charging sessions per day, distribution of arrival times on weekdays, distribution of arrival times on weekends, distribution of connection time per charging event, distribution of energy demand per charging event, development in distribution of EVs charing power are openly available. There maybe additional data sets available which needs download codes from ElaadNL.

3) What methods or software tools are needed to access the data? Is documentation about the software needed to access the data included? Is it possible to include the relevant software (e.g. in open source code)? Is there any documentation available on the dataset?

The datasets are available to download as .xls or .csv files or it can be also opened in highcharts cloud.

4) Assess the interoperability of your data. Specify what data and metadata vocabularies, standards, or methodologies you will follow to facilitate interoperability.

TBD

5) Specify how the dataset will be licensed to permit the widest reuse possible. What types of license do you plan to use?

Open.

6) Specify when the data will be made available for re-use. If applicable, specify why and for what period a data embargo is needed.

No data embargo is associated.

7) Specify whether the data produced and/or used in the project is useable by third parties, in particular after the end of the project. If the re-use of some data is restricted, explain why.

No restrictions. The Data Analytics team of ElaadNL is working in close collaboration with (inter)national research institutes on several aspects of electric mobility research.

2.1.4.3 Energy data (Netherland)

Dataset Name	*BAG – Basisregistratie Addressen en Gebouwen (Key register of addresses and buildings)*
Description	*The Key Register of Addresses and Buildings (BAG) is part of the Dutch government's system of key registers. Municipalities are owners of the BAG data. They are responsible for including the data in the BAG and for its quality. All municipalities make data about addresses and buildings available centrally via the National Facility BAG (LV BAG). The Land Registry manages the LV BAG and makes the data available to the various customers. Organizations with a public task, such as ministries, water boards, police forces and security regions are obliged to use the authentic data from the registrations.*

Security & Privacy considerations	*Public domain data*
Contact Person	*The Netherlands' Cadastre, Land Registry and Mapping Agency (dataplatform@kadaster.nl)*
Datatype Name	*Building details*
Description	*Year of construction, surface area, purpose and geographical co-ordinates of a building*
Purpose	*Data can be used for estimate of energy usage in a building*
Format	*JSON*
Expected Size	*Data size varies per area chosen for simulation*
Origin / Source	*BAG API Server (https://bag.bagregistraties.overheid.nl)*
Access	*Access free to public with API key*
Recipients	*CODEC-ESSIM simulation engine*
Metadata	*None*

Table 13 BAG

Dataset Name	*NEDU Energy Usage Profiles*
Description	*The Dutch Energy Data Exchange Association (NEDU) is the connecting platform for the Dutch energy sector. All market parties in the energy sector are represented here per market role. The various market roles work together in this to realize innovation and process improvement in the mutual processes. Market roles from NEDU make proposals for changes to be made to the regulations (information code) regarding the way in which these parties exchange data.*
Security & Privacy considerations	*Public domain data*
Contact Person	*De Vereniging Nederlandse EnergieDataUitwisseling (NEDU) (secretariaat@nedu.nl)*
Datatype Name	*Energy Usage Profiles*
Description	*Country-averaged electricity and gas usage profiles normalised over a year grouped into consumer types*
Purpose	*Energy use pattern per consumer can be estimated*
Format	*CSV*
Expected Size	*5.5MB for 2021 Dataset*
Origin / Source	*NEDU website (https://www.nedu.nl/documenten/verbruiksprofielen/)*
Access	*Access free to public*
Recipients	*CODEC-ESSIM simulation engine*
Metadata	*None*

Table 14 NEDU energy usage profiles

Dataset Name	*Service Sector and Urban-Scale Energy Demand*
Description	*The dataset contains demand profiles of 13 types of service sector consumers (hourly resolution, full year). - Demand profile of 1 type of average household consumer (hourly resolution, full year). - Demand profile of an average mix of 100 000 households and associated services, with a total annual demand of 710 GWh (hourly resolution,*

	full year). - Demand profile of 203 005 households only, also with a total annual demand of 710 GWh (hourly resolution, full year). - Demand profiles of archetype residential, business, and mixed urban areas. Urban areas include neighbourhoods, districts, and municipalities (hourly resolution, average weekday and average weekend). - Composition of archetype residential, business, and mixed urban areas. Urban areas include neighbourhoods, districts, and municipalities. - Spreadsheet tool to estimate the average hourly demand profile of an urban area of interest, based solely on annual demand data of different consumer types. All profiles pertain to the Netherlands and to the year 2014]
Security & Privacy considerations	*Public domain data*
Contact Person	*researchdata@4tu.nl*
Datatype Name	*Service Sector and Urban-Scale Energy Demand*
Description	*Service Sector and Urban-Scale Energy Demand*
Purpose	*Real urban areas consist of a mix of households, services (such as schools, offices, shops, etc.), and industry. However, most literature concerned with local energy demand simplifies it to household demand only. This is, to a large extent, cause by a lack of detailed (e.g., hourly) service sector and urban-scale energy demand data. This dataset and the accompanying thesis seek to resolve this issue*
Format	*.csv*
Expected Size	*9.8 MB*
Origin / Source	*4TU Research Data*
Access	*Public*
Recipients	*[UX researchers]*
Metadata	*None*

Table 15 Service sector and urban.scale energy demand

2.2 Operational and Observational Data

2.2.1 EUROSTAT indicators

Dataset Name	*EUROSTAT indicators*
Description	*Various EU, EEA, and beyond data for indicators such as GDP per capita, population, electricity consumption, etc.*
Security & Privacy considerations	*The data is downloaded from EUROSTAT and is aggregated at a country or NUTS level*
Contact Person	*a.iannone@cyberethicslab.com*
Datatype Name	*Macroeconomic indicators*
Description	*Used in descriptive analyses*
Purpose	*To contextualize the BRIGHT project*
Format	*Csv/xlsx*
Expected Size	*30KB*
Origin / Source	*EUROSTAT*
Access	*Open access*
Recipients	*WP3 partners*

Metadata	*No metadata created. All metadata is derived from EUROSTAT*

Table 16 EUROSTAT indicators

1) **How long will the data be stored/available?**

Data collected by CEL will be stored for 5 years after the end of the Project, and in any case in full compliance with Italian and EU regulations with respect to personal data storing.

2) **Which data will be made openly available? If some data is kept closed provide rationale for doing so.**

As a general rule, all the non-personal data will be openly available. As per personal data, they will be securely stored and shared with the Consortium or with some Partners upon necessity and in full compliance with GDPR and Italian data protection rules/laws.

3) **What methods or software tools are needed to access the data? Is documentation about the software needed to access the data included? Is it possible to include the relevant software (e.g. in open source code)? Is there any documentation available on the dataset?**

Excel or any relevant statistical software package.

4) **Assess the interoperability of your data. Specify what data and metadata vocabularies, standards, or methodologies you will follow to facilitate interoperability.**

In CSV the data is interoperable.

5) **Specify how the dataset will be licensed to permit the widest reuse possible. What types of license do you plan to use?**

Eurostat has a policy of encouraging free re-use of its data, both for non-commercial and commercial purposes. All statistical data, metadata, content of web pages or other dissemination tools, official publications and other documents published on its website, with the exceptions listed below, can be reused without any payment or written licence provided that:

- the source is indicated as Eurostat;
- when re-use involves modifications to the data or text, this must be stated clearly to the end user of the information.[1]

6) **Specify when the data will be made available for re-use. If applicable, specify why and for what period a data embargo is needed.**

The data is accessible online at links provided in deliverables. There is no embargo necessary.

7) **Specify whether the data produced and/or used in the project is useable by third parties, in particular after the end of the project. If the re-use of some data is restricted, explain why.**

No restriction necessary. The analysis of the data can be made available for up to 5 years after the end of the project.

2.2.2 Questionnaire responses from pilots

Dataset Name	*Questionnaire responses from pilots*
Description	*Response data from questionnaires submitted to pilot participants in WP3.*

[1] https://ec.europa.eu/eurostat/about/policies/copyright

Security & Privacy considerations	*Data will be held by CEL with state of the art solutions. Access to data will be provided only to CEL researchers strictly involved in the project*
Contact Person	*a.iannone@cyberethicslab.com*
Datatype Name	*Questionnaire responses from pilots*
Description	*Used in descriptive analyses*
Purpose	*To conduct activities in T3.3*
Format	*TBD* - Further details will be provided in the updated DMP (i.e. D1.4 – M18).
Expected Size	*TBD* - Further details will be provided in the updated DMP (i.e. D1.4 – M18).
Origin / Source	*BRIGHT Consortium*
Access	*Open access*
Recipients	*WP3 partners*
Metadata	*At this stage we do not envision applying existing metadata standards such as DDI. We will, however, note the date, time, and place of the data collected.*

Table 17 Questionnaire response from pilots

1) **How long will the data be stored/available?**

Data collected by CEL will be stored for 5 years after the end of the Project, and in any case in full compliance with Italian and EU regulations with respect to personal data storing.

2) **Which data will be made openly available? If some data is kept closed provide rationale for doing so.**

As a general rule, all the non-personal data will be openly available. As per personal data, they will be securely stored and shared with the Consortium or with some Partners upon necessity and in full compliance with GDPR and Italian data protection rules/laws.

3) **What methods or software tools are needed to access the data? Is documentation about the software needed to access the data included? Is it possible to include the relevant software (e.g. in open source code)? Is there any documentation available on the dataset?**

Excel or any relevant statistical software package. Further details will be provided in the updated DMP (i.e. D1.4 – M18).

4) **Assess the interoperability of your data. Specify what data and metadata vocabularies, standards, or methodologies you will follow to facilitate interoperability.**

We will exercise a best-effort policy to render the data interoperable. Further details will be provided in the updated DMP (i.e. D1.4 – M18).

5) **Specify how the dataset will be licensed to permit the widest reuse possible. What types of license do you plan to use?**

Creative Commons Licensing CC BY is foreseen to encourage maximum reuse, provided that the data is shared in anonymized form. This license lets others distribute, remix, adapt, and build upon your work, even commercially, as long as they credit you for the original creation. This is the

most accommodating of licenses offered. Recommended for maximum dissemination and use of licensed materials.[2]

6) Specify when the data will be made available for re-use. If applicable, specify why and for what period a data embargo is needed.

The analysis of data is planned to be accessible through public deliverables (i.e. D3.1, D3.2, D3.3 and D3.4), and potentially through publications as well. There is no embargo necessary. Further details will be provided in the updated DMP (i.e. D1.4 – M18).

7) Specify whether the data produced and/or used in the project is useable by third parties, in particular after the end of the project. If the re-use of some data is restricted, explain why.

No restriction necessary. The analysis of the data can be made available for up to 5 years after the end of the project.

2.2.3 Partner's answers to Data Protection Questionnaire

Dataset Name	*Partner's answers to Data Protection Questionnaire*
Description	*To comply with Project's Ethics Requirements a specific questionnaire concerning personal data processing activities was sent to all the Partners to provide their answers.*
Security & Privacy considerations	*Answers gathered are mostly non personal data; nevertheless CEL saved and stored the answers received on its own file sharing environment, ensuring the same degree of security as per CEL company's data. In addition, access to those data is granted on the need to know principle to the CEL personnel, which in any case is also bound by specific NDAs obligations.*
Contact Person	*a.iannone@cyberethicslab.com / e.sartini@cyberethicslab.com*
Datatype Name	*Data protection questionnaire answers*
Description	*Data gathered were used to prepare D10.1 and D10.2 as well as to map the flow of personal data processing within the perimeter of the Project*
Purpose	*To prepare and then submit D10.1 and D10.2, as well as to ensure Ethics compliance on the part of the Project and all its Partners, it was necessary to preliminarily understand to which extent Partners will process personal data in order to fulfill Projects objectives.*
Format	*MS Word DOCX*
Expected Size	<1MB
Origin / Source	*BRIGHT Consortium*
Access	Restricted to the Consortium and EC Services
Recipients	*CEL*
Metadata	*At this stage we do not envision applying existing metadata standards. We will, however, note the date, time.*

Table 18 Partner's answers to data protection questionnaire

1) How long will the data be stored/available?

[2] https://creativecommons.org/licenses/

Data collected by CEL will be stored for 5 years after the end of the Project, and in any case in full compliance with Italian and EU regulations with respect to personal data storing.

2) **Which data will be made openly available? If some data is kept closed provide rationale for doing so.**

As a general rule, all the non-personal data will be openly available. As per personal data, they will be securely stored and shared with the Consortium or with some Partners upon necessity and in full compliance with GDPR and Italian data protection rules/laws.

3) **What methods or software tools are needed to access the data? Is documentation about the software needed to access the data included? Is it possible to include the relevant software (e.g. in open source code)? Is there any documentation available on the dataset?**

There is no need to use a specific software.

4) **Assess the interoperability of your data. Specify what data and metadata vocabularies, standards, or methodologies you will follow to facilitate interoperability.**

We will exercise a best-effort policy to render the data interoperable. However, MS WORD DOCX is an interoperable format.

5) **Specify how the dataset will be licensed to permit the widest reuse possible. What types of license do you plan to use?**

This specific dataset is not reusable and will not be licensed for reuse.

6) **Specify when the data will be made available for re-use. If applicable, specify why and for what period a data embargo is needed.**

N/A

7) **Specify whether the data produced and/or used in the project is useable by third parties, in particular after the end of the project. If the re-use of some data is restricted, explain why.**

No, this kind of data cannot be used for other activities, especially from third parties not belonging to the Consortium.

As a final comment please consider that further details, comments as well potential new data sets can be included in the updated version of the DMP (D1.4) due at M18.

3 Findable, Accessible, Interoperable, and Reusable (FAIR) Data

3.1 Making data findable, including provisions for metadata

To make data generated/collected by the project findable, it is important to have a naming and versioning convention. For Project documents, this convention has been already defined in the Quality Management Plan. Deliverables and other Project documents must be circulated among partners following this naming convention:

BRIGHT_DX.Y_Vk.j_PPP

where:

- DX.Y: is the deliverable number according to the Grant Agreement;
- Vk.j: is the version number, V1.0 is the final version to be sent to the Commission;
- PPP: is the partner's abbreviation responsible for a specific version of the document. The value of this field for the latest version of the document (V1.0) is the name of the partner responsible for the deliverable.

For example, document with title "BRIGHT_D1.0_V1.0_ENG" indicates Final version (V1.0) of the deliverable D1.0 which is delivered by partner Engineering.

In addition to the naming convention, the "Guidelines on FAIR Data Management in Horizon 2020" also propose to have Digital Object Identifiers (DOIs) for the data generated during the project. BRIGHT will use the Zenodo platform (http://www.zenodo.org) for the data that will be made publicly available. Zenodo is an open-access repository developed under the European OpenAIRE program and operated by CERN that also fulfils the DOI requirement. The platform can manage single datasets with up to 50GB size. To help research projects to share data all over the world, the platform also helps by defining and storing some additional metadata provided by the uploader. It is possible to grant access to the data only to a specific group of users or the public. The platform also gives the user the possibility to restrict or open access to data for a fixed period of time.
Regarding the provision of metadata, BRIGHT will not use any formal standards for its creation. Instead, security and privacy guidelines will be followed to ensure that only the necessary details are stored, especially when dealing with sensitive personal information. Each dataset will include its own list of meaningful metadata.

3.2 Making data openly accessible

An important aspect of FAIR data management is to make the data accessible to project partners and, when possible, to external parties such as other researchers and the public. BRIGHT will use the Zenodo platform for the data that the Consortium will decide to make public.
Due to the nature of the data acquired during the pilot implementations, not all data collected will be made openly available. Some of the information acquired is private data subject to GDPR and national regulations and therefore cannot be opened without reservation. All collected data, such as real-time energy consumption and renewable energy production, before being made available will be anonymized or aggregated and will not be used with any identifying information within the project or in external publications.

3.3 Making data interoperable

Making data interoperable mainly relies on using suitable standards for data and metadata creation along with appropriate vocabularies (e.g., for providing search keywords).
To facilitate interoperability a partial list of standards used by BRIGHT datasets includes:

- Standard ontology like SAREF to create reference language for energy-related data;
- OpenADR 2.0 communication protocol;
- Secure TCP/IP and MQTT protocols to communicate with the IoT devices and energy assets;
- RESTful API services to communicate among the different software components;
- CSV and JSON formats data.

Other standards may be added in future updated versions of the document.

3.4 Increase data re-use (through clarifying licences)

Data will be stored in servers, either on site of the pilots or on locations indicated by the technology provider. Public data will be published after the release of the respective deliverable or after the end of the project. The availability of data that contain key information on the end customer's commercial operations should be discussed with pilot sites partners following the terms of the consortium agreement and with the consent of the end user. However, sensitive data will be anonymized and processed/analysed as a part of a larger body of data. No information, from which an individual participant can be identified, will be published. Only anonymized results will be summarized as a part of a research publication.
The availability of data after the end of the project depends highly on the type and content of the data. Therefore, storing data on a public platform needs to be discussed with the contributor of the data. In general, data will be made publicly available on a half-yearly to yearly basis during the project lifetime.
Regarding the application of licenses, Creative Commons Licensing CC BY is foreseen to encourage maximum reuse, provided that the data is shared in anonymized form.

4 Personal data protection's principles and Management of Personal Data

The activities of processing of personal data follow specific rules and are subject to certain laws and regulations. Consequently, with reference to the protection of personal data, the Project, as well as each Partner, is committed to fully comply with all the applicable laws and regulations. In particular each Partner is fully aware on its duties and obligations as set forth within the EU General Data Protection Regulation 2016/679 ("**GDPR**").

To this extent, and having in mind that the protection of personal data and privacy rights entails the protection of two fundamental rights recognized at EU level and in each Member States of the Union, the respect of GDPR and of other applicable laws shall be ensure vis-à-vis personal data belonging to either individuals not belonging to the Consortium and to those belonging to the same.

It is therefore worth briefly summarizing some of the most important principles applicable in case of data processing.

First of all, personal data shall be processed lawfully, fairly and in a transparent manner in relation to the data subject. To this extent, each Partner commits to process personal data in a way that ensures compliance with the principle of purpose limitation, data minimization, accuracy, storage limitation, integrity, confidentiality and accountability.

In light of the above, in any case of processing of personal data, the interested Partner will remain accountable and responsible for the data collected during the Project and shall be liable to identify the most appropriate lawful basis before starting any processing operations. For those processing activities for which consent (ex art. 6. 1. a) of GDPR) is the most appropriate legal basis, the relevant Partner will make reference to Annex II – BRIGHT Privacy Notice and Consent Form of D10.2.

Personal data shall be processed in a manner that ensures appropriate security, including protection against unauthorized or unlawful processing and against accidental loss, destruction or damage, using appropriate technical or organizational measures.

Technical or organizational measures shall permit identification of data subjects for no longer than is necessary for the purposes for which the personal data are processed. At the end of the Project, all processed personal data shall be destroyed in accordance with the specific law requirements.

Against this backdrop, in order to understand if (i) Partners are going to process, jointly or severally, personal data for the purposes of the Project and (ii) which is the flow and management of the said personal data, between M1 and M2 of the Project (i.e. November and December 2020) CEL submitted to all the Partners a data protection questionnaire ("**Data Protection Questionnaire**").

According to the answers directly provided by the Partners the deliverables pertaining to WP 10 – Ethics Requirements were submitted to the EC, as well as it resulted that it is possible that for certain activities Partners will process personal data. However, in consideration to the fact that the Data Protection Questionnaire was submitted at a very early stage of the Project, CEL decided to submit again the following questions to have a clearer picture now at M6 of the Project (i.e. April 2021):

"*For the purposes of the Project,*

- *are you going to process personal data?*
    - *If yes, or if you are not sure, for which activities?*

- *If yes, or if you are not sure, are you going to share these personal data inside the Consortium?*
- *If yes, or if you are not sure are you going to share these personal data outside the Consortium?*

- *Are you going to processe personal data falling within the definition of article 9 of GDPR? (i.e. personal data revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, or trade union membership, and the processing of genetic data, biometric data for the purpose of uniquely identifying a natural person, data concerning health or data concerning a natural person's sex life or sexual orientation).*
  - *If yes, or if you are not sure, for which activities?*
  - *If yes, or if you are not sure, are you going to share these personal data inside the Consortium?*
  - *If yes, or if you are not sure, are you going to share these personal data outside the Consortium?*"

4.1 Day-to Day Data Usage and Processes related to Project Management

In light of the above, the present paragraph is dedicated to identifying those set of activities entailing personal data processing of individuals belonging to the BRIGHT Consortium.

Scenario	Solution
BRIGHT mailing lists	For the purposes of achieving BRIGHT results, and to manage the work flow among all the Partners involved, it has been created a series of mailing lists, namely (i) a general one, where have been indicated at least one contact person per Partner, and (ii) one mailing list for each of the work packages. In particular, the scope of the lists is to keep updated the relevant Partners upon tasks, events, and the progress of the Project in general. The mailing lists created are restricted only to BRIGHT Partners, and the end of the Project will be erased. As per the management of such mailing lists, i.e. the additions and/removals, are responsibility of the Coordinator (ENG). In any case, each person included in a mailing has the right to opt-out by contacting the Coordinator.
Meetings and related material	During BRIGHT meetings (either virtual or in person) it is possible that documents will be created and used, such as agendas, presentations, minutes and signature lists etc. These documents will be created and managed only inside the Consortium and its Partners, and will be used only for the purposes of the relevant meeting. Moreover, each Partner might have access to the document, which to this extent will be stored in Project's shared environment. The storage of these documents will be limited to 5 years after the end of the Project. To the extent permissible by law, any person whose personal data will be included therein shall have the right to request at any time to the Coordinator to opt-out.
Workshops/Conferences, training and dissemination sessions	Events such as workshops, conferences, and plenary meetings might be attended by one or more individuals belonging to the Consortium. In this scenario, personal data such as name, surname, company affiliation, emails, and pictures/video recording might be

	collected. Such data might be collected and processed not only for the purposes of organising the said event, but also for dissemination. In the latter case, before the publication, the relevant individual might request to opt-out from the publications by emailing to the Coordinator. The data will be stored in the BRIGHT shared environment, and the data will be kept for 5 years after the Project.
Reporting	Reports providing for updates on the Project progress, as well as on financial data, might contain personal data. These reports might be shared either within and outside the Consortium for compliance purposes with national financial law, and in particular with the EC.
Deliverables, internal documents, and other BRIGHT reports	During the lifetime of the Project, a large series of documents and reports will be produced, like deliverables and/or internal documents etc. These files will be used to fulfil Project contractual obligations and shared to: BRIGHT Partners, EC, and, depending on the nature of the document, shared with external individuals (as this might be the case for those deliverables that are classified as public and that might be published on the project's website). In these documents, the name and/or email of authors may be included. As far as the said documents are going to be shared inside the Consortium and distributed to the EC, they will be used only for the purposes of reporting and stored in the BRIGHT cloud server under the deliverables section. Reports that will be shared publicly (public deliverables) will mention only the Partner name and not any other personal information, unless agreed otherwise with the relevant author. All documents will be kept for 5 more years after the Project ends.
Other scenarios not included above	As a general principle, in any case according to which kind personal data needs to be added in any kind of document for the purposes of the Project, the controller (i.e. the document creator) shall have to notify the data subject that his/her personal data will be included into the related document, specifying purposes, retention period, storage requirements etc.

5 Conclusions and next steps

This deliverable reports the first version of the BRIGHT Data Management Plan. The main goal of this version was to identify the datasets used for the project, including all the related information available at this stage.

The document will be updated whenever significant changes arise, for example when new datasets are introduced or consortium policies change. Two additional versions of this document are already planned at M18 (April 2022) D1.4 and at M36 (October 2023) D1.6.

Later versions may include additional information on a finer granularity about the dataset already descibed or describe more in detail aspects not mentioned in the current version, such as data retention policies.