The PARITY project has received funding from the EU's Horizon 2020
research and innovation programme under grant agreement No 864319

Project Acronym:	**PARITY**
Project Full Title:	**Pro-sumer AwaRe, Transactive Markets for Valorization of Distributed flexibilITY enabled by Smart Energy Contracts**
Grant Agreement:	**846319**
Project Duration:	**42 months (01/10/2019 – 31/03/2023)**

DELIVERABLE D2.2
Data Management Plan

Work Package:	**WP2 – Quality Assurance and Management**
Task:	**T2.1 – Governance, technical coordination and quality assurance**
Document Status:	**Final v3.0**
File Name:	**PARITY_D2.2_Data Management Plan_R1_V3.0_CERTH**
Due Date:	**March 2022**
Submission Date:	**June 2022**
Lead Beneficiary:	**CERTH**

Dissemination Level	
Public	X
Confidential, only for members of the Consortium (including the Commission Services)	

The PARITY project has received funding from the EU's Horizon 2020 research and innovation programme under grant agreement No 864319

Authors List

Leading Author				
First Name		**Last Name**	**Beneficiary**	**Contact e-mail**
Dimosthenis		Ioannidis	CERTH	djoannid@iti.gr
Co-Author(s)				
#	**First Name**	**Last Name**	**Beneficiary**	**Contact e-mail**
1	Stylianos	Zikos	CERTH	szikos@iti.gr

Reviewers List

Reviewers			
First Name	**Last Name**	**Beneficiary**	**Contact e-mail**
All partners	All partners	All partners	-

Version History

Version	Author	Date	Status
0.1	Dimosthenis Ioannidis, Stylianos Zikos, CERTH	January 10, 2020	Initial draft (TOC)
0.4	Stylianos Zikos, CERTH	February 3, 2020	Added initial content
0.6	Stylianos Zikos, CERTH	March 2, 2020	Added information about datasets
0.9	Dimosthenis Ioannidis, Stylianos Zikos, CERTH	March 10, 2020	Final draft for internal review
1.0	Dimosthenis Ioannidis, Stylianos Zikos, CERTH	March 30, 2020	Final version (v1) including comments from partners, ready for submission
2.0	Stylianos Zikos, CERTH	March 29, 2021	Final version (v2) with updates on the datasets, ready for submission
3.0	Stylianos Zikos, CERTH	May 30, 2022	Final version (v3) with updates on the datasets, ready for submission

Legal Disclaimer

The PARITY project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 864319. 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.

Copyright

Executive Summary

The present document, titled as “Data Management Plan”, describes the data management framework and is a detailed report for the open research data of the project. Regarding the data management framework, general principles are presented, including data handling and open access policy. Furthermore, data management processes to be applied during and after the end of the project are described, such as data collection, storage, retention and destruction.

In the last part of the document, the different types of data to be collected within the project’s activities and pilot use cases are presented, as well as various datasets that have been utilised and produced. Regarding the datasets, a dataset template was created by taking into account the “Guidelines on FAIR Data Management in Horizon 2020” and considering the feedback received from partners. The template was sent to partners responsible for the development of specific components or data collection from participants, in order to collect information about the anticipated datasets. Main information included refers to the description of the dataset and its relation to project’s objectives, the sharing policy, performed pre-processing steps, its format and other properties. A dedicated community has been created on Zenodo open repository in order to host the open datasets of PARITY, facilitating re-use and exploitation.

The Data Management Plan has been a living document; this is the third and final version of the document that includes the updated descriptions of all datasets that have been created, shared or planned to be shared.

Table of Contents

List of Figures

List of Tables

List of Acronyms and Abbreviations

Term	Description
API	Application Programming Interface
CSV	Comma Separated Values
DER	Distributed Energy Resources
DMP	Data Management Plan
DPO	Data Protection Officer
EC	European Commission
EU	European Union
EV	Electric Vehicle
FAIR	Findable Accessible Interoperable Re-usable
GDPR	General Data Protection Regulation
IML	Information Management Layer
IPR	Intellectual Property Rights
JSON	JavaScript Object Notation
KPI	Key Performance Indicator
LEM	Local Energy Market
LFM	Local Flexibility Market
NDA	Non-Disclosure Agreement
OA	Open Access
ORDP	Open Research Data Pilot
P2H	Power-to-heat
PIR	Passive Infrared
SFTP	Secure File Transfer Protocol
SLA	Service Level Agreement
SSL	Secure Sockets Layer
SVN	Subversion
URI	Uniform Resource Identifier
URL	Uniform Resource Locator
V2G	Vehicle to Grid
WP	Work Package
XML	Extensible Markup Language

1. INTRODUCTION

1.1 Scope and Objectives of the Deliverable

This document presents the Data Management Plan (DMP) that has been defined and will be applied within the PARITY project. The principles, processes and standards that are described in this report are going to be taken into account by PARITY partners during the collection, storage, processing and sharing of the generated datasets related to project's activities.

The main objective of the document is to provide information on the types of data the project will generate, how they will be exploited or made accessible for verification and re-use, and how it will be curated and preserved. The goal is to have open the majority of research data generated by the project and make them available through the Open Research Data Pilot (ORDP). ORDP aims to improve and maximise access to and re-use of research data generated by Horizon 2020 projects. It takes into account several aspects, such as protection of information, commercialisation, Intellectual Property Rights (IPR) and privacy & security [1].

Different types of datasets are presented in the Data Management Plan. The first type concerns datasets that are derived after processing the data collected from surveys and interviews involving participants. Another dataset type concerns raw data from pilot sites' sensors and devices (e.g. smart meter data). Moreover, datasets derived from the output data of software components that will be developed within the project will also be generated. Open datasets can be used by third parties for further research and validation, after a post-processing phase.

1.2 Structure of the Deliverable

The document is structured as follows: Section 2 presents the plan for data management. Information on several topics is provided, such as general principles of the DMP, and procedures for data description, data security, data collection and storage. Furthermore, it presents the data identification form template that has been created. Section 3 includes a list of data that will be collected and descriptions of the identified datasets within PARITY project. Lastly, conclusions are provided in Section 4.

1.3 Relation to Other Tasks and Deliverables

This deliverable is an outcome of *Task 2.1 - Governance, technical coordination and quality assurance* included in *WP2 - Quality Assurance and Management*. The deliverable is related to *Task 1.1 – Ethics Management*, as the DMP has been developed by taking into account the ethical issues documented in *D1.1 – Ethics Manual*, which was the outcome of T1.1.

Information provided in this document can be used as input for the following tasks:

- Tasks related to exploitation activities, such as *Task 10.3 – Business innovation planning for project results* and *Task 10.5 – PARITY scaling-up and replication roadmap*.
- Tasks related to research and standardization recommendations, such as *Task 9.4 – Establishment of synergies and coordination with BRIDGE and similar projects for policy and research-relevant issues* and *Task 9.5 – Grid integration and standardization recommendations for LFMs*.
- Tasks related to the development of PARITY components within *WP5 – Local Flexibility Market Platform*, *WP6 – Smart Grid Optimization & Management*, and *WP7 – DER Flexibility Management & Storage-as-a-Service*.

2. DATA MANAGEMENT FRAMEWORK

This section describes the PARITY data management framework, which defines the format of the datasets and handles issues related to data such as:

- data handling and sharing
- data protection
- data collection and storage
- data retention and destruction

2.1 General Principles

Data Management Plans (DMPs) [2] are a key element of good data management. A 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 re-usable (FAIR), a DMP should include information on:

- the handling of research data during and after the end of the project
- what data will be collected, processed and/or generated
- which methodology and standards will be applied
- whether data will be shared
- how data will be curated and preserved

PARITY participates in the *Pilot on Open Research Data* launched by the EC. Data that will be generated within the project activities may be available in open access for processing and exploitation. All relevant procedures for data handling are described in the DMP in order to ensure the proper usage of data.

2.1.1 Data Availability and Handling

The term Open Access (OA) refers to a set of principles and practices through which scientific results are distributed online, free of charge or other access barriers. The benefits of broader access to scientific publications and data are mentioned in the "*Guidelines to the Rules on Open Access to Scientific Publications and Open Access to Research Data in Horizon 2020*" [3] and are the following:

- Improved quality of results, as it is possible to be built on previous research results
- Greater efficiency since duplication is avoided and collaboration is enhanced
- Faster progress to market and faster growth
- Improved transparency of the scientific process due to higher involvement of citizen and society

Regarding the availability of data, three different categories are defined as showed in Table 1.

Table 1. Data categories regarding availability.

Data categories	Description
Public	This category refers to data that are publicly shared for re-use and exploitation
Consortium	This category refers to data that are available only to the members of the project and the EU Commission services and are subject to Non-Disclosure Agreement.
Confidential	This category refers to data that are retained by individual partner of the project.

Data handling is another important issue for the project. Data handling activities will be carried out by the data controllers and the data processors, in compliance with the General Data Protection Regulation (GDPR), and as described in *D1.1 – Ethics Manual* [4]. Data controllers determine the purposes and means of the processing of personal data, while the data processors process the personal data on behalf of the controller. Within PARITY, pilot responsible partners and the Data Protection Officers (DPOs) will act as the data controllers, whereas the technical partners who will analyze the data will have the data processor role. For the datasets that will be shared among the consortium partners in order to accomplish the objectives of the tasks, NDAs are foreseen to be signed among involved data processors and data controllers prior to distribution of the data.

2.1.2 Open Access to Scientific Publications

As already mentioned, Open Access refers to the practice of providing online access to scientific information that is free of charge to the end-user and reusable [3]. The term 'scientific information' in the context of research and innovation can refer to either:

- peer-reviewed scientific research articles published in scholarly journals, or
- research data which can be data underlying publications, curated data and/or raw data.

There are two main options for open access that will be considered:

Self-archiving or 'green' open access – the author, or a representative, archives (deposits) the published article or the final peer-reviewed manuscript in an online repository, such as zenodo.org, before, at the same time as, or after publication. Some publishers request that open access be granted only after an embargo period has elapsed.

Open access publishing or 'gold' open access - an article is immediately published in open access mode. In this model, the payment of publication costs is shifted away from subscribing readers.

The most common business model is based on one-off payments by authors. These costs, often referred to as article processing charges are usually borne by the researcher's university or research institute or the agency funding the research. In other cases, the costs of open access publishing are covered by subsidies or other funding models.

2.1.3 Open Access to Research Data

Open Access to Research Data refers to the right to access and reuse digital research data under the terms and conditions set out in the Grant Agreement. In a research context, examples of data include statistics, results of experiments, measurements, observations resulting from fieldwork, survey results, interview recordings and images. The focus is on research data that is available in digital form. Users can normally access, mine, exploit, reproduce and disseminate openly accessible research data free of

charge. Figure 1 visualizes the different data dissemination and exploitation options in a flowchart diagram.

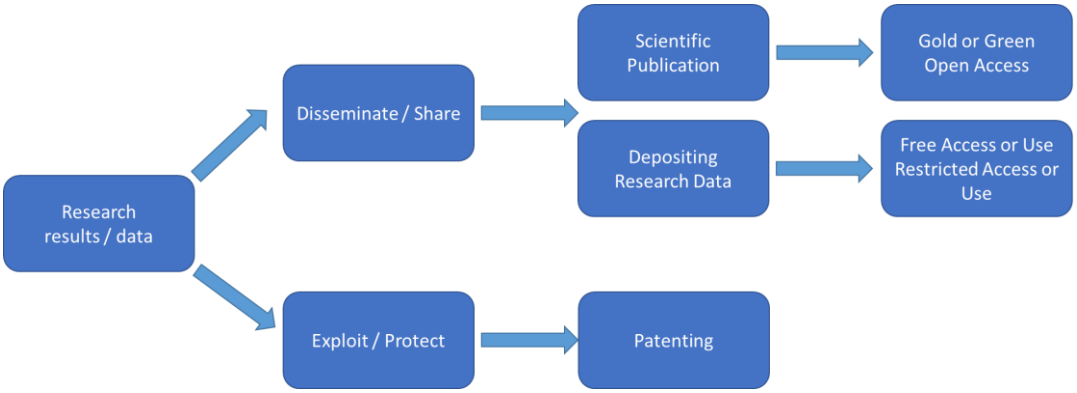

Figure 1. Data dissemination and exploitation options.

The four main characteristics of open data are denoted by the acronym FAIR [2]:

Findable: To facilitate discovery and identification, data have a unique ID, located in a searchable resource, and documented with meaningful metadata.

Accessible: Data are readily and freely retrievable using common methods and protocols. Metadata are accessible even if the data are not.

Interoperable: Data are presented in broadly recognized standard formats, vocabularies, and languages, in order to facilitate sharing and analysis.

Re-useable: Data are provided under a clear license. Accurate meaningful metadata conform to relevant community standards and identifying its content and provenance.

2.2 Data Protection and Security

Ethical or legal issues that can have an impact on data protection and sharing may exist as data from end users' facilities and equipment will be collected and processed. A security by design approach is followed during the project in order to reduce the security and privacy risks. A privacy and security framework has been developed within *T3.4 – PARITY Privacy & Security Framework Specifications*. Initially the security risks were identified and the security and privacy requirements of the pilot owners have been collected and taken into account.

In order to protect the collected data and control unauthorised access to the PARITY data repositories, only authenticated personnel will have access to use case-specific data collected. The use of secure protocols such as SSL will help to protect privacy. Towards the protection of personal data of volunteer research participants, the following issues will be taken into account:

- All personal data will be held private;
- Personal data on subjects will be used in strictly confidential terms and will only be published as statistics (anonymously);
- Personal data will be held private, regardless of how this data was acquired. Therefore, data obtained incidentally within project will be handled with confidentiality. This accidental obtainment does not substitute the compulsory procedure, in which researchers need each participant's explicit consent to obtain, store and use information about them;

- Personal data will be pseudo-anonymised in full and at the earliest possible point in time during data processing;
- The acquired personal data will under no circumstances be used for commercial purposes.
- Where re-identification is needed in order to perform the forensic tasks, the traceability of the origin of the personal data will be removed and the relevant information required to perform such task will be kept separately and only for such purposes.
- In cases where additional information about end users is captured, the captured data will be also pseudo-anonymised.

The issues related to ethics, data protection and security as well as the measures that will be taken to overcome them, are presented in *D1.1 Ethics Manual* [4]. All the procedures will be followed in compliance with EU and national legislation and regulations, including GDPR. Basic personal data, such as name, age, and contact information, may need to be collected. In this case, the consortium will protect such data in accordance with the EU Data Protection Directive 95/46/EC concerning the protection of individuals with regard to the processing of personal data and on the free movement of such data.

2.3 Data Collection and Storage

Collected data will be anonymized and stored in secure servers, only in case it is required for performing the project's tasks. Re-identification will be only possible to be performed by specific personnel from pilot partners, technical manager, and project coordinator.

2.3.1 Data Collection

Certain actions will be performed, in order to ensure proper data collection:

- Data transfer from installed devices to IoT Gateway and repositories will be automated and secure.
- Devices, and developed blockchain and IoT software that will be installed at pilot sites will ensure secure data collection from energy meters.
- Secure transfer protocols will be considered, such as SFTP, for both automated and manual data transfers.
- Data aggregation methods will be applied on raw data of individual prosumers, so that aggregators, market operators and DSOs will not be aware of the data of individual prosumer.

2.3.2 Data Storage

Regarding user data that will be stored, such as energy consumption and generation, usage patterns, thermal comfort profiles, etc., a pre-processing step will be applied for anonymization and aggregation prior to storage. User identification data may be stored and used only for problem solving during the development phase and the pre-validation phase. All data that will be considered confidential will be erased as soon as they are not needed anymore or at the end of the project. On the contrary, data that will be characterized as public will be kept open to the research community along with all accompanied metadata information. These data will be available on Zenodo, at the PARITY community that has been created, which is accessible at the URL [5].

2.4 Data Retention and Destruction

A number of critical factors which are relevant for data retention will be taken into account, namely:

- Purpose of retaining data
- Type of open data collected
- Policy access to the open data
- Data storage, security and protection measures

- Confidentiality and anonymity of data

Regarding data destruction, as computerized data (hard disk drives) will be used for data storage, existing methods for permanent and irreversible destruction of the data will be utilized, such as full disk overwriting and re-formatting tools. In all cases, the data protection and privacy of personal data will be governed by the following principles, which consist of part of an overall information security policy:

- Protective measures against infiltration will be provided
- Physical protection of core parts of the systems and access control measures will be provided
- Logging of Project system and appropriate auditing of the peripheral components will be available

2.5 Data Sharing

PARITY facilitates safe data sharing by fully enforcing and protecting user rights. For the datasets that will be shared only among the consortium members, authorization and authentication mechanisms are applied in order to restrict access and ensure that only the appropriate users can retrieve the data. Datasets will be provided in the form of files, such as .csv, in the most common case. However, Application Programming Interfaces (APIs) have been developed for exchanging data among different PARITY components. These APIs can be utilized by users to retrieve part or all the data in the appropriate format, e.g. XML or JSON. APIs that act on public data may also be provided to third parties in case no other option is possible.

In summary, all activities related to data management will be performed in full compliance with any European and national legislation and directives relevant to the country where the data collections are taking place (international or European): *The Universal Declaration of Human Rights and the Convention 108 for the Protection of Individuals with Regard to Automatic Processing of Personal Data*; and *GDPR & Directive 2002/58/EC of the European parliament regarding issues with privacy and protection of personal data and the free movement of such data*. Moreover, respective data authorities will be informed or their opinion will be requested, in case respective experimental/piloting activities (within WP8) raise ethics, privacy, security or any other implications (i.e. misuse).

2.6 Datasets Format

Each dataset will be described by the following properties included in the data identification form (or dataset template), that is shown in Table 2. The dataset template was created by taking into account the Guidelines on FAIR Data Management in Horizon 2020 [2] and includes various sections: Data Identification, Related WPs and Tasks, Partners Responsibilities, Exploitation, Metadata, Pre-processing, Sharing and Expected Size.

Table 2. Dataset template.

<DS-XX-Title>	
Data Identification	
Dataset name	<Short name/outline of the dataset>
Dataset description	<Explain why and how data are collected and describe the content. Also, mention the related datasets (used as input), if any.>

	Related datasets:
Origin of the data (e.g. device, evaluation surveys)	<Mention how the dataset will be collected>
Related PARITY component(s)	<Mention the related PARITY component>
Relation to the PARITY objectives	<Mention the related PARITY objectives>
Related WPs and Tasks	
WPs	<The data will be collected within activities of WPX>
Tasks	<The data will be collected within activities of tasks TX>
Partners Responsibilities	
Partner(s) responsible for the data collection	<Partner short name>
Partner(s) responsible for the data storage	<Partner short name>
Partner(s) responsible for the data analysis	<Partner short name>
Metadata, Pre-processing, Sharing and Expected Size	
Metadata and documentation	<Provide information about the discoverability of the dataset and if a documentation will be available>
External data used	<Mention the external data/database sources in case the dataset includes data from sources outside of PARITY project>
Data pre-processing steps	<Mention any pre-processing done to the dataset: E.g. anonymization, data interpolation, outlier cleaning etc.>
Sharing	<Public / Consortium / Confidential>
Licence type (e.g. Public Domain, Attribution, Non-commercial, No Derivatives, or other)	<Mention the type of licence>
Expected volume of data	<Mention the expected size of data>
Format of data	<Mention the format of data, e.g. JSON, csv etc.>
Storage location (URI)	<Mention where dataset is stored: e.g. public URL, SVN path>
Exploitation	
Data utility (to whom will it be useful)	<Explain shortly the purpose of the data collection>

3.DATA TO BE COLLECTED AND DATASETS

Within PARITY, the majority of data will be collected automatically by sensors and other devices installed at the pilot sites, or generated by PARITY services or components. Additionally, other data will be collected directly from end users through questionnaires and interviews at the workshops that will be organised. Examples of data in the former case are the following:

- Energy consumption and production data from DERs
- Energy network related information and network modelling data
- Energy demand and production forecasted data
- Prosumer flexibility forecasted data
- Data from sensors (temperature, humidity, occupancy etc.)
- User control actions over controllable devices (e.g., manual change of HVAC system setpoint)
- User comfort profiles
- EV charging schedules
- Billing and transactions data through the blockchain platform

3.1 Datasets List

The list of datasets that is shown in Table 3 has been derived from the functionalities of the modules and description of objectives & tasks, as presented in the Description of Action document of the project [6]. The datasets are related to activities performed within WP3 - Use Cases, Requirements and System Architecture Definition, WP4 - Local Flexibility Business & Market Models, WP6 - Smart Grid Optimization & Management, WP7 - DER Flexibility Management & Storage-as-a-Service and WP10 – Exploitation and Business Innovation.

Table 3. Datasets List.

Dataset ID-Name	Related Tasks
DS-01-Requirements questionnaires results	T3.1
DS-02-A-Prioritization of barriers that hinders Local Flexibility Market proliferation	T4.1
DS-02-B-Barriers and drivers detection: Taxonomy of barriers that hinders Local Flexibility Market proliferation	T3.1, T4.1
DS-03-Network modelling data	T6.1, T6.2, T6.4, T7.3, T7.4, T7.5, T10.2
DS-04-Indoor Ambient Conditions	T7.1
DS-05-Metering Data	T7.1
DS-06-Occupancy Presence	T7.1
DS-07-User control actions	T7.1
DS-08-EV Charging Schedules (V1G, V2G)	T7.3
DS-09-Assets Clusters	T7.5
DS-10-E-mobility urbener	T7.3
DS-11-Smart contracts and citizens	T4.2

DS-12-Energy market transformation and the customers' perspective questionnaire results	T10.1
DS-13-Stationary battery status data	T7.4, T7.5

3.2 Description of Datasets

This section presents the characteristics of the identified or already generated datasets, based on the PARITY dataset template. Different data collection methods have been applied to create the described datasets, and the sharing properties of each dataset is in accordance with its scope and exploitation.

DS-01-Requirements questionnaires results	
Data Identification	
Dataset name	Requirements questionnaires results
Dataset description	This dataset will provide inputs from different stakeholders to define the main requirements and characteristics of the PARITY tool suite. *Related datasets: No other datasets*
Origin of the data (e.g. device, evaluation surveys)	Surveys and interviews
Related PARITY component(s)	Mainly requirements and characteristics of the PARITY tool suite to be developed
Relation to the PARITY objectives	Understanding incentives and barriers for flexibility adoption
Related WPs and Tasks	
WPs	WP3, WP4, WP5, WP6 and WP7
Tasks	Task 3.1, other task as the main requirements of the PARITY tool suite are analysed.
Partners Responsibilities	
Partner(s) responsible for the data collection	CIRCE
Partner(s) responsible for the data storage	CIRCE
Partner(s) responsible for the data analysis	CIRCE
Metadata, Pre-processing, Sharing and Expected Size	
Metadata and documentation	None
External data used	None
Data pre-processing steps	Anonymization
Sharing	Consortium (Only processed data, not the original)
Licence type (e.g. Public Domain, Attribution, Non-commercial, No Derivatives, or other)	PD

Expected volume of data	1 MB – 5MB
Format of data	xls, csv, word, voice
Storage location (URI)	CIRCE TEAMS software and storage servers
Exploitation	
Data utility (to whom will it be useful)	PARITY project partners to know the main requirements and characteristics of the PARITY tool suite.

DS-02-A-Prioritization of barriers that hinders Local Flexibility Market proliferation	
Data Identification	
Dataset name	Prioritization of barriers that hinders Local Flexibility Market proliferation
Dataset description	This dataset contains the prioritization provided by a panel of 15 experts to a set of 28 barriers categories for 8 different roles of the future energy system. A Delphi method was followed and the scores provided in the three rounds carried out are included. The dataset also contains the scripts used to assess the results and the output of this assessment. *Related datasets: No other datasets*
Origin of the data (e.g. device, evaluation surveys)	Delphi Method, surveys + focus groups
Related PARITY component(s)	Use cases, barriers taxonomy
Relation to the PARITY objectives	Understanding incentives and barriers for flexibility adoption
Related WPs and Tasks	
WPs	WP4
Tasks	T4.1
Partners Responsibilities	
Partner(s) responsible for the data collection	DEUSTO
Partner(s) responsible for the data storage	DEUSTO
Partner(s) responsible for the data analysis	DEUSTO
Metadata, Pre-processing, Sharing and Expected Size	
Metadata and documentation	It will be published in Zenodo repository when conveniently anonymised

External data used	None
Data pre-processing steps	Anonymization
Sharing	Public
Licence type (e.g. Public Domain, Attribution, Non-commercial, No Derivatives, or other)	Creative Commons Attribution 4.0 International
Expected volume of data	less than 1 MB
Format of data	csv
Storage location (URI)	https://doi.org/10.5281/zenodo.3861847
Exploitation	
Data utility (to whom will it be useful)	Campaigners and intervention designers that want to understand the potential barriers that different stakeholders may encounter when addressing energy flexibility measures and objectives.

DS-02-B-Taxonomy of barriers that hinders Local Flexibility Market proliferation	
Data Identification	
Dataset name	Taxonomy of barriers that hinders Local Flexibility Market proliferation
Dataset description	This dataset contains the results of the state of the art survey to retrieve the barriers that hinders Local Flexibility Market proliferation. Scientific literature, interviews with different stakeholders, technical reports from the main energy agencies and the European and national legislation have been consulted to build the taxonomy. *Related datasets: No other datasets*
Origin of the data (e.g. device, evaluation surveys)	Surveys + Interviews
Related PARITY component(s)	Stakeholders and roles
Relation to the PARITY objectives	Understanding incentives and barriers for flexibility adoption
Related WPs and Tasks	
WPs	WP3 and WP4
Tasks	T3.1 and T4.1
Partners Responsibilities	
Partner(s) responsible for the data collection	DEUSTO, CIRCE
Partner(s) responsible for the data storage	DEUSTO
Partner(s) responsible for the data analysis	DEUSTO
Metadata, Pre-processing, Sharing and Expected Size	
Metadata and documentation	It will be published in Zenodo repository when conveniently anonymised
External data used	None
Data pre-processing steps	Anonymization
Sharing	Public
Licence type (e.g. Public Domain, Attribution, Non-commercial, No Derivatives, or other)	Creative Commons Attribution 4.0 International
Expected volume of data	less than 1 MB
Format of data	ods
Storage location (URI)	https://doi.org/10.5281/zenodo.3863018
Exploitation	
Data utility (to whom will it be useful)	Campaigners and intervention designers that want to understand the potential barriers that different stakeholders may encounter when addressing energy flexibility measures and objectives.

DS-03-Network modelling data	
Data Identification	
Dataset name	Network modelling data
Dataset description	These data provide the necessary information for the modelling of the low voltage networks (topology, generation profiles, consumption profiles, voltage profiles...). *Related datasets: No other datasets*
Origin of the data (e.g. device, evaluation surveys)	Demo partners
Related PARITY component(s)	Mainly requirements and characteristics of the PARITY tool suite to be developed.
Relation to the PARITY objectives	Understanding incentives and barriers for flexibility adoption
Related WPs and Tasks	
WPs	WP6 and WP7
Tasks	Task 6.1, Task 6.2, Task 6.4, Task 7.3, Task 7.4, Task 7.5 and Task 10.2
Partners Responsibilities	
Partner(s) responsible for the data collection	CIRCE
Partner(s) responsible for the data storage	CIRCE
Partner(s) responsible for the data analysis	CIRCE
Metadata, Pre-processing, Sharing and Expected Size	
Metadata and documentation	None
External data used	None
Data pre-processing steps	Anonymization
Sharing	Consortium (Only processed data, not the original)
Licence type (e.g. Public Domain, Attribution, Non-commercial, No Derivatives, or other)	PD
Expected volume of data	1 MB – 5MB
Format of data	xls, csv, pfd
Storage location (URI)	CIRCE TEAMS software and storage servers
Exploitation	
Data utility (to whom will it be useful)	PARITY project partners to know the main requirements and characteristics of the PARITY tool suite. Moreover, part of the resulting information will be used for the validation of business models in T10.2.

DS-04-Indoor Ambient Conditions	
Data Identification	
Dataset name	Indoor Ambient Conditions
Dataset description	Indoor ambient conditions collected from sensor devices installed inside the building. Data are transferred through the IoT Gateway where a JSON message of collected data is created. *Related datasets:* Temperature *(^oC), Humidity (%), Illuminance (Lux)*
Origin of the data (e.g. device, evaluation surveys)	Sensor devices, IoT Gateway
Related PARITY component(s)	PARITY Oracle, Information Management Layer (IML), P2H component
Relation to the PARITY objectives	Relation to all the objectives described in WP7
Related WPs and Tasks	
WPs	The data will be collected within activities of WP7
Tasks	The data will be collected within activities of task T7.1
Partners Responsibilities	
Partner(s) responsible for the data collection	Hypertech
Partner(s) responsible for the data storage	Hypertech
Partner(s) responsible for the data analysis	Hypertech
Metadata, Pre-processing, Sharing and Expected Size	
Metadata and documentation	Those data will not be made available
External data used	None
Data pre-processing steps	Data cleansing techniques (e.g. outlier detection/removal/replacement)
Sharing	Confidential
Licence type (e.g. Public Domain, Attribution, Non-commercial, No Derivatives, or other)	Non-applicable
Expected volume of data	The expected volume of data is about 600MB per zone and per year
Format of data	Timeseries in JSON format
Storage location (URI)	Hypertech's IML cloud server
Exploitation	
Data utility (to whom will it be useful)	Comfort profiling engine, Demand and Flexibility Control/Optimization engine (P2H models, BaaB module), SLAs management

DS-05-Metering Data	
Data Identification	
Dataset name	Metering Data
Dataset description	Metering data for power consumption will be collected from smart meters, power meters and smart clamps. These measurements can refer to specific devices that will to the aggregated consumption of a building zone. *Related datasets:* Power (W), Energy (Wh), Voltage (V), Current (A)
Origin of the data (e.g. device, evaluation surveys)	Metering data are collected by Power Meters, Smart Clamps, Smart Plugs and Nano Switches, HIVE pre-existing smart meters
Related PARITY component(s)	IoT Gateway, Information Management Layer (IML), P2H component and BaaB component
Relation to the PARITY objectives	Relation to all the objectives described in WP7
Related WPs and Tasks	
WPs	The data will be collected within activities of WP7
Tasks	The data will be collected within activities of tasks T7.1
Partners Responsibilities	
Partner(s) responsible for the data collection	Hypertech, CheckWatt, HIVE
Partner(s) responsible for the data storage	Hypertech, CheckWatt, HIVE
Partner(s) responsible for the data analysis	Hypertech, e7, CIRCE
Metadata, Pre-processing, Sharing and Expected Size	
Metadata and documentation	The data will not be made publicly available
External data used	None
Data pre-processing steps	Data cleansing procedure including outlier detection and/or replace/removal
Sharing	Consortium
Licence type (e.g. Public Domain, Attribution, Non-commercial, No Derivatives, or other)	Non-applicable
Expected volume of data	The expected volume of data is about 600MB per zone and per year
Format of data	Timeseries in JSON format
Storage location (URI)	Hypertech's IML Cloud Server, BigQuery database, CheckWatt database/server
Exploitation	

Data utility (to whom will it be useful)	Demand and Flexibility Control/Optimization engine (P2H model, BaaB module), SLAs management and KPI validation; part of the data will be used for the evaluation of business models in T10.2

DS-06-Occupancy Presence	
Data Identification	
Dataset name	Occupancy Presence
Dataset description	Identify occupancy inside a building zone based on several motion detections from one or a group of motion sensors *Related datasets: Motion (binary values)*
Origin of the data (e.g. device, evaluation surveys)	Occupancy data are collected by PIR motion sensors
Related PARITY component(s)	PARITY Oracle, Information Management Layer (IML), Occupancy profiling module (sub-component of P2H component)
Relation to the PARITY objectives	Relation to all the objectives described in WP7
Related WPs and Tasks	
WPs	The data will be collected within activities of WP7
Tasks	The data will be collected within activities of tasks T7.1
Partners Responsibilities	
Partner(s) responsible for the data collection	Hypertech, CERTH
Partner(s) responsible for the data storage	Hypertech, CERTH
Partner(s) responsible for the data analysis	CERTH
Metadata, Pre-processing, Sharing and Expected Size	
Metadata and documentation	The data will not be made publicly available
External data used	None
Data pre-processing steps	No pre-processing steps is expected to be utilized
Sharing	Consortium
Licence type (e.g. Public Domain, Attribution, Non-commercial, No Derivatives, or other)	Non-applicable
Expected volume of data	The expected volume of data is about 200MB per zone and per year
Format of data	JSON
Storage location (URI)	Hypertech's IML cloud server, LEM/LFM

	Repository
Exploitation	
Data utility (to whom will it be useful)	Occupancy profiling engine, Comfort profiling engine, Demand and Flexibility Control/Optimization engine (P2H and BaaB components), SLAs management

DS-07-User control actions	
Data Identification	
Dataset name	User control actions
Dataset description	These data include the user control actions and configuration in several type of devices like HVAC, DHW and Lighting devices. *Related datasets: Status (On/Off), Setpoint (°C), Mode (Heating/Cooling), Brightness (%)*
Origin of the data (e.g. device, evaluation surveys)	User control actions data are originally collected by smart device controllers
Related PARITY component(s)	PARITY Oracle, Information Management Layer (IML), P2H component
Relation to the PARITY objectives	Relation to all the objectives described in WP7
Related WPs and Tasks	
WPs	The data will be collected within activities of WP7
Tasks	The data will be collected within activities of tasks T7.1
Partners Responsibilities	
Partner(s) responsible for the data collection	Hypertech
Partner(s) responsible for the data storage	Hypertech
Partner(s) responsible for the data analysis	Hypertech
Metadata, Pre-processing, Sharing and Expected Size	
Metadata and documentation	The data will not be made publicly available
External data used	None
Data pre-processing steps	Data cleansing techniques (e.g. outlier detection/removal/replacement)
Sharing	Confidential
Licence type (e.g. Public Domain, Attribution, Non-commercial, No Derivatives, or other)	Non-applicable
Expected volume of data	The expected volume of data is about 600MB per zone and per year

Format of data	JSON
Storage location (URI)	Hypertech's IML cloud
Exploitation	
Data utility (to whom will it be useful)	Comfort profiling engine, Demand and Flexibility Control/Optimization engine (P2H component), SLAs management

DS-08-EV Charging Schedules (V1G, V2G)	
Data Identification	
Dataset name	EV Charging Schedules (V1G, V2G)
Dataset description	EV charging schedules will be generated automatically by the PARITY EV Profiling and Smart Charging software component based on dynamic electricity pricing, EV characteristics and other parameters *Related datasets: No other datasets*
Origin of the data (e.g. device, evaluation surveys)	PARITY EV Profiling and Smart Charging software component will generate the data
Related PARITY component(s)	EV Profiling and Smart Charging component, LEM/LFM Repository
Relation to the PARITY objectives	Exploit flexibility of EV storage capacity, Optimise EV charging
Related WPs and Tasks	
WPs	The data will be collected within activities of WP7
Tasks	The data will be collected within activities of task T7.3
Partners Responsibilities	
Partner(s) responsible for the data collection	CERTH, BFS, URBENER
Partner(s) responsible for the data storage	BFS, URBENER, CERTH
Partner(s) responsible for the data analysis	CERTH, SUPSI
Metadata, Pre-processing, Sharing and Expected Size	
Metadata and documentation	Documentation and additional information is available online
External data used	EV charging requests from pilot site
Data pre-processing steps	Anonymization
Sharing	Consortium / Public

Licence type (e.g. Public Domain, Attribution, Non-commercial, No Derivatives, or other)	Creative Commons Attribution 4.0 International
Expected volume of data	Up to 1MB per day depending on number of EVs and charging stations
Format of data	csv, xlsx files
Storage location (URI)	https://doi.org/10.5281/zenodo.6461328
Exploitation	
Data utility (to whom will it be useful)	Output schedules will be used as input to T7.5 flexibility management tools; partly to be used also in T10.2 for the evaluation of the business models.

DS-09-Assets Clusters	
Data Identification	
Dataset name	Assets Clusters
Dataset description	Aggregator Toolset will generate this dataset in order to contribute to higher flexibility utilization *Related datasets: EV V2G Schedules, Building as a Battery services datasets*
Origin of the data (e.g. device, evaluation surveys)	PARITY software component (service) will generate the data
Related PARITY component(s)	Aggregator Toolset (Prosumer Flexibility Manager, VPP Manager)
Relation to the PARITY objectives	Maximize benefit through flexibility exchange and LFM participation
Related WPs and Tasks	
WPs	The data will be collected within activities of WP7
Tasks	The data will be collected within activities of task T7.5
Partners Responsibilities	
Partner(s) responsible for the data collection	CERTH, Pilot responsible partners
Partner(s) responsible for the data storage	CERTH, Pilot responsible partners
Partner(s) responsible for the data analysis	CERTH, Technical partners
Metadata, Pre-processing, Sharing and Expected Size	
Metadata and documentation	This information will be available
External data used	None
Data pre-processing steps	Anonymization
Sharing	Consortium
Licence type (e.g. Public Domain, Attribution, Non-commercial, No Derivatives, or other)	Attribution Non-commercial
Expected volume of data	0.5 MB - 5 MB (depending on the number of assets and prosumers)
Format of data	JSON
Storage location (URI)	PARITY SVN repository
Exploitation	
Data utility (to whom will it be useful)	Data will be useful to technical partners for performing evaluation, and to prosumers for better understanding participation to markets through aggregation.

DS-10-E-mobility urbener	
Data Identification	
Dataset name	E-mobility urbener
Dataset description	Control energy consumption, location, identification and date of the EV charge *Related datasets:* Energy consumption (kWh)
Origin of the data (e.g. device, evaluation surveys)	Energy consumption data are collected by Urbener software
Related PARITY component(s)	EV energy consumption
Relation to the PARITY objectives	Predict and assess the untapped flexibility of EV to optimally integrate them into the PARITY concept
Related WPs and Tasks	
WPs	The data will be collected within activities of WP7
Tasks	The data will be collected within activities of task 7.3
Partners Responsibilities	
Partner(s) responsible for the data collection	URBENER
Partner(s) responsible for the data storage	URBENER
Partner(s) responsible for the data analysis	URBENER
Metadata, Pre-processing, Sharing and Expected Size	
Metadata and documentation	A dataset document will be available with the consumption in kW and the time (duration) in minutes that it has been charging.
External data used	None
Data pre-processing steps	Anonymization
Sharing	Consortium
Licence type (e.g. Public Domain, Attribution, Non-commercial, No Derivatives, or other)	Public
Expected volume of data	219 kB
Format of data	xlsx
Storage location (URI)	URBENER PLATFORM
Exploitation	
Data utility (to whom will it be useful)	Data will be useful for optimization of EV charges; partly to be used also in T10.2 for the evaluation of the business models

DS-11-Smart contracts and citizens	
Data Identification	
Dataset name	Smart contracts and citizens
Dataset description	This dataset was collected in the context of T4.2 and deliverable *D4.2 Design of next-generation smart-contract-enabled energy contracts*. The dataset consists of the results of an anonymous survey conducted through Prolific platform (prolific.co) to elucidate the users' perceptions about automatic trading systems. 832 responses were collected and coded. The organization of the information collected was: 1. A short survey to get perceptions from a scenario describing an automatic trading system. 2. Demographic information.
Origin of the data (e.g. device, evaluation surveys)	Surveys: Prolific
Related PARITY component(s)	Smart contracts. Blockchain.
Relation to the PARITY objectives	PARITY aims at delivering a market for automated flexibility exchange based on smart contracts & blockchain
Related WPs and Tasks	
WPs	WP4
Tasks	Task 4.2
Partners Responsibilities	
Partner(s) responsible for the data collection	UDEUSTO
Partner(s) responsible for the data storage	UDEUSTO
Partner(s) responsible for the data analysis	UDEUSTO
Metadata, Pre-processing, Sharing and Expected Size	
Metadata and documentation	It will be published in Zenodo
External data used	None
Data pre-processing steps	Cleaning
Sharing	Public
Licence type (e.g. Public Domain, Attribution, Non-commercial, No Derivatives, or other)	PD
Expected volume of data	less than 1 MB
Format of data	csv
Storage location (URI)	https://doi.org/10.5281/zenodo.4558723
Exploitation	

Data utility (to whom will it be useful)	Anyone who might be interested in confronting citizens to automated systems, emerging technologies and the adoption of blockchain based technologies.

DS-12-Energy market transformation and the customers' perspective survey results	
Data Identification	
Dataset name	Energy market transformation and the customers' perspective survey results
Dataset description	This dataset will provide inputs from the PARITY experts, stakeholders and end-users, to explore the overall patterns that will form the energy market transformation and the consumer's perspective. *Related datasets: No other datasets*
Origin of the data (e.g. device, evaluation surveys)	Survey
Related PARITY component(s)	The dataset is related to all the PARITY tools and services, providing them insights related to the market needs.
Relation to the PARITY objectives	Exploring the patterns that form the energy market transformation, measuring awareness regarding energy market and gaining insights on energy market tools and services.
Related WPs and Tasks	
WPs	The data were collected within activities of WP10
Tasks	The data were collected within activities of T10.1
Partners Responsibilities	
Partner(s) responsible for the data collection	UNIC
Partner(s) responsible for the data storage	UNIC
Partner(s) responsible for the data analysis	UNIC
Metadata, Pre-processing, Sharing and Expected Size	
Metadata and documentation	n/a
External data used	None
Data pre-processing steps	Anonymization
Sharing	Consortium
Licence type (e.g. Public Domain, Attribution, Non-commercial, No Derivatives, or other)	Public Domain

Expected volume of data	About 20 kB
Format of data	.xlsx
Storage location (URI)	• PARITY SVN repository • https://doi.org/10.5281/zenodo.6420882
Exploitation	
Data utility (to whom will it be useful)	PARITY project partners to gain market insights related to the PARITY tool suite, anyone who is interested to gain knowledge regarding the progress of local renewable energy development and in motivation behind customers' participation in such markets.

DS-13-Stationary battery status data	
Data Identification	
Dataset name	Stationary battery status data
Dataset description	The dataset contains the values of six different variables of the stationary battery that is installed at the Smart House DIH of CERTH. These data are collected and stored automatically every minute or every 15 minutes (depending on the type of the variable).
Origin of the data (e.g. device, evaluation surveys)	Battery Energy Management System
Related PARITY component(s)	Building-as-a-Battery framework
Relation to the PARITY objectives	Relation to all the objectives described in WP7
Related WPs and Tasks	
WPs	The data are collected within activities of WP7
Tasks	The data are collected within activities of T7.4 and T7.5
Partners Responsibilities	
Partner(s) responsible for the data collection	CERTH
Partner(s) responsible for the data storage	CERTH
Partner(s) responsible for the data analysis	CERTH
Metadata, Pre-processing, Sharing and Expected Size	
Metadata and documentation	Short description available
External data used	None
Data pre-processing steps	Days with missing data points were excluded
Sharing	Consortium
Licence type (e.g. Public Domain, Attribution, Non-commercial, No Derivatives, or other)	Attribution Non-commercial
Expected volume of data	2 MB
Format of data	csv
Storage location (URI)	PARITY SVN repository /Datasets/Stationary_Battery_Certh_Smarthouse DIH
Exploitation	
Data utility (to whom will it be useful)	The data can be useful for analysing the operation of an actual stationary battery and testing energy management applications.

4. CONCLUSIONS

This document presented the principles and procedures included in the PARITY data management plan, which is important to be followed for proper exploitation of research data. Main topics that are handled by the data management plan are the open access policy, data handling and sharing, and procedures that will be followed for data collection, storage, retention and destruction. These procedures will facilitate the protection of user rights, as well as the work of the consortium.

In addition to the data management framework, the different types of data that have been collected were presented along with their relation to project's tasks and objectives. Particular attention has been given to making datasets publicly available where possible. Those datasets have been made available online at the PARITY community on Zenodo, which is the open repository developed under the European OpenAIRE program. Furthermore, it shall be noted that data pre-processing steps were applied in several cases in order to ensure high quality of provided data as well as users' privacy. The first version of this document provided descriptions of the initially identified datasets. Dataset descriptions are provided with the use of a dataset template, which was created by taking into account the Guidelines on FAIR Data Management in Horizon 2020. Descriptions of three new datasets were added, while other two datasets were updated in the second version of the document. The current version of the document, which is the final version, provided a presentation of the final datasets and their properties up to March 2022. Lastly, any additional open datasets that will be generated in the next months until the end of the project, will be published on the project's repository in Zenodo along with their descriptions.

5.REFERENCES

[1]. Guidelines on Implementation of Open Access to Scientific Publications and Research Data, version 1.1, 21 April 2017, https://ec.europa.eu/research/participants/data/ref/h2020/other/hi/oa-pilot/h2020-hi-erc-oa-guide_en.pdf

[2]. European Commission, Guidelines on FAIR Data Management in Horizon 2020, version 3.0, 26 July 2016, https://ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/open-access-data-management/data-management_en.htm

[3]. European Commission, Guidelines to the Rules on Open Access to Scientific Publications and Open Access to Research Data in Horizon 2020, version 3.2, 21 March 2017, https://ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/open-access-data-management/open-access_en.htm

[4]. D1.1 Ethics Manual, PARITY project, Public Deliverable, 2020

[5]. Zenodo PARITY community, https://zenodo.org/communities/parity-h2020

[6]. The PARITY EU research project Description of Action, GA Number 864319