CBDC powered Smart PerFORrmance contracTs for Efficiency, Sustainable, Inclusive, Energy use

D1.2 Data management plan. Report and updates

Report Identifier:	D1.2		
Work-package:	WP1	**Task:**	T1.5
Responsible Partner:	National Technical University of Athens (NTUA)	**Version Number:**	1.0
Due Date	31/03/2023	**Document Date**	31/03/2023
Distribution Security:	PU	**Deliverable Type:**	R
Keywords:	Data management, FAIR data, IPR management, Datasets		
Project website: http://www.fortesie.eu/			

Quality Control

	Name	Organisation	Date
Editor	Christos Kontzinos	NTUA	22/03/2023
Peer review 1	Sonia García	CTIC	27/03/2023
Peer review 2	Vasilis Stavrakas	IEECP	29/03/2023
Authorised by **(Technical Coordinator)**	Alkiviadis Giannakoulias	ED	30/03/2023
Authorised by **(Quality Manager)**	Kostas Panagopoulos	ED	30/03/2023
Submitted by **(Project Coordinator)**	Anastasia Garbi	ED	31/03/2023

Legal Disclaimer

FORTESIE is an EU project funded by the Horizon Europe (HORIZON) research and innovation programme under grant agreement No. 101080029. The information and views set out in this deliverable are those of the author(s) and do not necessarily reflect the official opinion of the European Union. The information in this document is provided "as is", and no guarantee or warranty is given that the information is fit for any specific purpose. Neither the European Union institutions and bodies nor any person acting on their behalf may be held responsible for the use which may be made of the information contained therein. The FORTESIE Consortium members shall have no liability for damages of any kind including without limitation direct, special, indirect, or consequential damages that may result from the use of these materials subject to any liability which is mandatory due to applicable law.

Table of Contents

List of Figures

List of Tables

Abbreviations

AHU	Air Handling Unit
AIoD	Artificial Intelligence on Demand
API	Annual Percentage Rate
APR	Application Programming Interface
BIPV	Building Integrated PhotoVoltaics
CBDC	Central Bank Digital Currency
DB	Database
DMP	Data Management Plan
DPIA	Data Protection Impact Assessment
EC	European Commission
EPC	Energy Performance Contract
ESIE	Efficient, Sustainable, and Inclusive Energy
EUCI	European Union Classified Information
FAIR	Findable, Accessible, Interoperable, Re-usable
GDPR	General Data Protection Regulation
GE	Green Euro
IDS	International Data Spaces
IoT	Internet of Things
IPMVP	International Performance Measurement and Verification Protocol
IPR	Intellectual Property Rights
JSON	JavaScript Object Notation
KYC	Know Your Customer
NGSI	Next Generation Service Interfaces
OA	Open Access

PDL	Power Distribution Limit
PV	PhotoVoltaics
SLA	Service Level Agreement
SSH	Social Sciences and Humanities
SSL/TLS	Secure Sockets Layer/Transport Layer Security
VFD	Variable Frequency Drive
XML	Extensive Markup Language

Executive Summary

This deliverable is part of FORTESIE WP1 – Project Management and specifically, it represents the work conducted for T1.5 – Data Management. Following the Horizon Europe Programme Guidelines in FAIR Data Management it constitutes the FORTESIE Data Management Plan (DMP) addressing the following issues:

- What is the data management life cycle for all datasets to be collected or generated and processed by the FORTESIE project?
- How will the project results and research data be handled during and after the project?
- What are the data that will be collected, processed, or generated?
- What methodology and standards will be applied?
- What is the data sharing policy?
- What processes will be followed for data curation and preservation?

Based on the "Guidelines on Data Management in Horizon 2020", the Data Management Plan aims to produce data so that researchers may benefit by their use directly, and/or to apply their methods based on data generated by Research in Horizon Europe. Such information would be the scientific publications issued by the project consortium, white papers published, Open-Source code generated, mock-up datasets used for supporting the development process etc. The Data Management Plan governs all data generated and collected during the project, the standards that will be used, how the research data will be preserved and what parts of the datasets will be shared for verification or reuse.

1 Introduction

Since data management is at the core of the FORTESIE project, the consortium will follow a series of dedicated activities in publishing, disseminating, spreading, and communicating the project data (outcomes and accumulated knowledge) to external parties, such as interested communities and potential stakeholders, with the aim to leverage existing and create new opportunities.

The main goals of this report are the following:

- To detail the overall methodology for handling the outcomes of the project, in accordance with the H2020 guidelines regarding Open Research Data.
- To list results, information and data that can be published.
- To describe the open repositories for data management and dissemination.

FORTESIE project partners will provide, through open access, various types of information, such as scientific publications relevant to the project, white papers published, Open-Source code generated, open datasets, anonymous interview results, etc. It should be stressed that the consortium will balance between open publishing project's related data, collected, or generated, and protecting private or sensitive information (according to GDPR provisions) that may have legal implications in case of inappropriate treatment.

1.1 Project Introduction

The overall vision of FORTESIE is to design, demonstrate, validate and replicate innovative renovation packages in the building industry with Smart Performance-Based guarantees and financing, aiming at Efficient, Sustainable and Inclusive Energy (ESIE) use to accelerate the Renovation Wave in Europe. The renovation packages will combine state-of-the-art construction materials and technologies components (prefabricated facades, BIPV, heat pumps, etc.), innovative digital technologies for measurement and verification, and attractive financing (e.g., contractual frameworks for smart performance guarantees, financing mechanisms, engagement techniques, green-euros, etc.), to raise the overall EPC (Energy Performance Contract) value proposition. The renovation packages will be tailored to specific target groups needs and optimised to improve ESIE performance considering energy, CO2, and comfort. Each package will be demonstrated and validated in real-life use cases and customised for replication in all other partner countries for immediate market take-up. Methodologies from Social Sciences and Humanities (SSH) will be adopted for: a. the creation of collaborative business models that boost the Renovation Wave by considering all stakeholders' value and revenue streams, b. novel incentivisation and behavioural change models that aim to stimulate long-term engagement with focused interactions to adopt green behaviour c. the incorporation of a digital currency, green-euro, (€G) for financing, rewarding and creating an inclusive / collective narrative in the fight against climate change d. the collection of feedback for recommendations to policy and business stakeholders, and e. the mapping and understanding the complex interplay between the different stakeholders to deliver an engagement strategy across the value chain. These demonstrations will potentially constitute the green euro as a retail Central Bank Digital Currency (CBDC), hence revolutionising the financing of renovation approaches. An online marketplace will be offering first level advice, directing consumers through the value chain of stakeholders and facilitating access to these "packaged" renovation services.

1.2 Deliverable Purpose

Based on the guidelines of the Open Research Data Pilot in Horizon 2020[1], the Data Management Strategy for FORTESIE will be based on the establishment of how data will be handled during and after the project. This process involves:

- A methodology that can make research data generated in the context of the FORTESIE project: findable, accessible, interoperable, and reusable (FAIR principles).
- The identification, classification (i.e., open, or confidential), organisation, curation, preservation, storing and sharing of the data to be collected, processed and/or generated.
- Requirements related to ethics and legal compliance (i.e., to ensure that the work will be conducted in an ethically sound way) as described in the Grant Agreement and in EU and national legislation.

1.3 Data Protection Legislative Framework

The FORTESIE consortium is fully aware of the ethical implications of the proposed research and respects the ethical rules and standards of Horizon Europe, and those reflected in the Charter of Fundamental Rights of the European Union. Where necessary, the FORTESIE consortium confirms its abidance to national and international laws including Regulation (EU) 2016/679[2] of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, the Directive on Privacy and Electronic Communications (2002/58/EC)[3], Directive on Protection of Privacy in the Telecommunication Sector (97/66/EC)[4], The Universal Declaration of Human Rights[5] and the Convention 108 for the Protection of Individuals with Regard to Automatic Processing of Personal Data[6]. Article 19 "Ethical principles" of Regulation No. 1291/2013/EC of the European Parliament[7] and of the Council which states the fundamental principles of the H2020 Ethics in research.

1.4 Structure of the Document

This deliverable is structured as follows:

- **Section 1** provides the introduction of the deliverable.
- **Section 2** presents the FORTESIE data management strategy, thereby exposing classification, archiving, performance, safety and security, FAIR and ethics requirements and procedures for the data.
- **Section 3** lists datasets identified at the early phases of the FORTESIE project.
- **Section 4** presents specific implementation aspects of the Data Management Strategy
- **Section 5** provides the summary and conclusions of the deliverable.

[1] http://ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/open-access-data-management/data-management_en.htm

[2] https://eur-lex.europa.eu/eli/reg/2016/679/oj

[3] https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=CELEX%3A32002L0058

[4] https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A31997L0066

[5] https://www.un.org/en/about-us/universal-declaration-of-human-rights

[6] https://www.coe.int/en/web/data-protection/convention108-and-protocol

[7] https://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2013:347:0104:0173:EN:PDF

2 Data Management Strategy and Procedures

Data Management Plans (DMPs) 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 by a Horizon Europe project. As part of making research data findable, accessible, interoperable, and reusable, a DMP should include information about the handling of research data during and after the end of the project:

- What (kind of) data will be collected, processed and/or generated and to whom might they be useful later on?
- Which methodology and standards will be applied?
- What metadata will be required to enable data to be found and understood, ideally according to the standards of a scientific discipline?
- Whether data will be shared/made open access.
- How data will be preserved (including after the end of the project)?
- How to archive and preserve the open datasets of the project?

More specifically, for Horizon Europe projects a FAIR DMP template[8] has been designed to be applicable to any project that produces, collects, or processes research data (please see Annex A). The FAIR data principles towards promptly disseminating the data outcomes of a research project[9] can be seen below in Table 2-1: Data Management according to the FAIR principles data source and acquisition.

Table 2-1: Data Management according to the FAIR principles data source and acquisition[9]

FAIR Data Principles	
Data should be **Findable**	F1. (Meta)data are assigned a globally unique and persistent identifier. F2. Data are described with rich metadata (defined by R1 below). F3. Metadata clearly and explicitly include the identifier of the data they describe. F4. (Meta)data are registered or indexed in a searchable resource.
Data should be **Accessible**	A1. (Meta)data are retrievable by their identifier using a standardised communication protocol. A1.1 The protocol is open, free, and universally implementable. A1.2 The protocol allows for an authentication and authorisation procedure, where necessary. A2. Metadata are accessible, even when the data are no longer available.
Data should be **Interoperable**	I1. (Meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.

[8] Guidelines on FAIR Data Management in Horizon 2020, http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf

[9] https://www.go-fair.org/fair-principles/

	I2. (Meta)data use vocabularies and definitions that follow FAIR principles. I3. (Meta)data include qualified references to other (meta)data.
Data should be **Reusable**	R1. (Meta)data are richly described with a plurality of accurate and relevant attributes. R1.1. (Meta)data are released with a clear and accessible data usage license. R1.2. (Meta)data are associated with detailed provenance. R1.3. (Meta)data meet domain-relevant community standards.

2.1 Data Sources and Acquisition

Data collected in FORTESIE are both public/open data available on the internet and internal operational data collected or generated from/by partners, mainly pilot and research organisations. The data collected in FORTESIE involves the following data sources:

- Document-based data, including:
  - Interviews and surveys with stakeholders participating in the pilots during requirements elicitation, as well as validation of the FORTESIE solution.
- Operational data produced during the project execution:
  - Public/Open data, such as energy consumption, weather, ESIE performance data gathered from sensors monitoring the buildings, etc.
  - Internal energy-related data from pilots, such as energy monitoring data, smart meters data, PV plant production, etc.
  - Mobile app (users' data) including user's profile, home description, and input regarding the behaviour model recommendations/tips, to be defined in more detail during project specification.
  - Data from the FORTESIE dissemination activities (e.g., website visitors and other website analytics, webinars participants)

2.2 Types of Data

In addition, a main point of the DMP is the definition of the open access type over the data. Open Access (OA) refers to the practice of providing online access to scientific information that is free of charge to the end-user and reusable. 'Scientific' refers to all academic disciplines. In the context of research and innovation, 'scientific information' can mean:

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

Open Access[10] to scientific publications means free online access for any user. The two main routes to Open Access are:

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

[10] Guidelines to the Rules on Open Access to Scientific Publications and Open Access to Research Data in Horizon 2020, https://ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/open-access-data-management/data-management_en.htm

- Open Access publishing / '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.

Research data refers to information, in particular facts or numbers, collected to be examined and considered as a basis for reasoning, discussion, or calculation. 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**: Open Access strategy for publications and research data presents the process flow towards defining the open access type in scientific publications and research data.

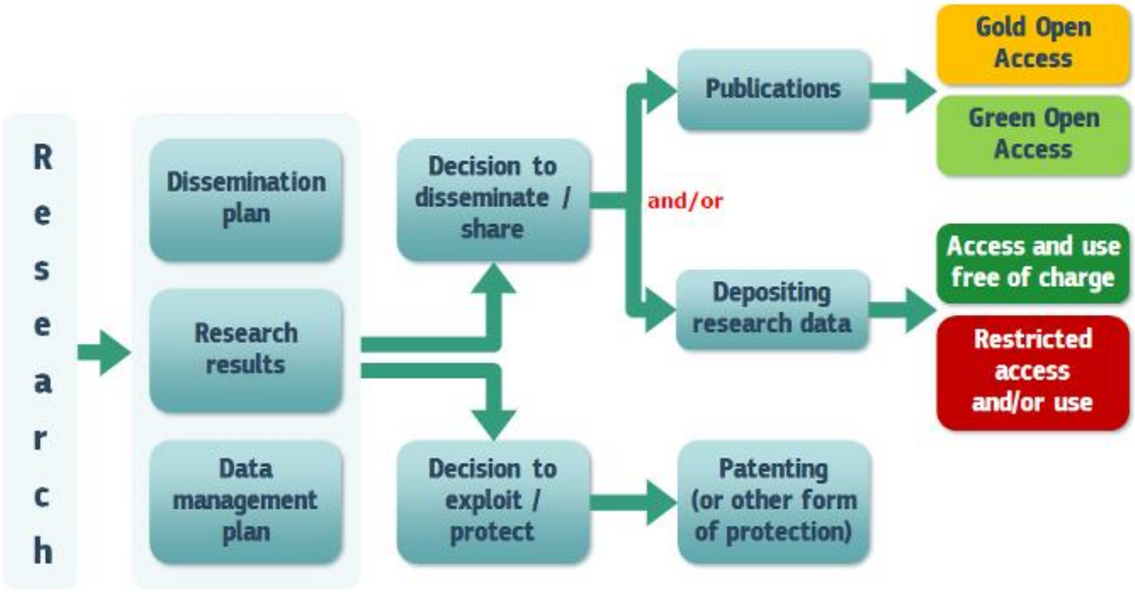

Research
Dissemination plan
Research results
Data management plan
Decision to disseminate / share
Decision to exploit / protect
and/or
Publications
Depositing research data
Patenting (or other form of protection)
Gold Open Access
Green Open Access
Access and use free of charge
Restricted access and/or use

Figure 1: Open Access strategy for publications and research data[11]

The open access mandate comprises two steps:

1. depositing publications in repositories
2. providing open access to them

2.3 Data Management Requirements

The FORTESIE Data Management Process is defined as a management approach for each result generated or collected during the project runtime. Such an approach outlines requirements for several data management aspects, including data classification, data archiving, data performance, data safety and security, FAIR data management and data ethics. These aspects refer to the FORTESIE platform as well as to other assets of the project, namely the FORTESIE Portal, the project's shared repository (ProofHub), the research data online repository (Zenodo), FORTESIE's code repository (GitHub), and the project communication channels. The FORTESIE requirements and corresponding

[11] https://ec.europa.eu/research/participants/docs/h2020-funding-guide/imgs/open-access.png

processes for the aforementioned data management aspects are presented in the following paragraphs.

2.3.1 Data Classification

For proper data handling, datasets ought to be primarily classified as public and non-public. The following questions must be answered to classify the different datasets:

1. **Does a result provide significant value to others or is it necessary to understand a scientific conclusion?**

If this question is answered with yes, then the result is classified as public (granted for open access). If this question is answered with no, the result is classified as non-public. For example, code that is very specific to the FORTESIE platform (e.g., a database initialisation) is usually of no scientific interest to anyone, nor does it add any significant contribution.

2. **Does a result include personal information that is not the author's name?**

If this question is answered with yes, the result is classified as non-public. Personal information beyond the name must be removed if it should be published according to the ethical principles of the project.

3. **Does a result allow the identification of individuals even without their names?**

This is also a step managed by the legal/ethical framework of the project as we have committed in the FORTESIE project to establish encryption techniques and store personal data securely. Datasets will be anonymised for impact assessment and research purposes. The personal data collected as part of the project will be limited to the project submission and informed consent of participants about the use of personal data will be required. Personal identity will be protected by the use of anonymous codes. If this question is answered with yes, the result is classified as non-public.

4. **Can a result be abused for a purpose that is undesired by society in general or contradicts societal norms and the project's ethics?**

If this question is answered with yes, the result is classified as non-public.

5. **Does a result include business or trade secrets of one or more partners of the project?**

If this question is answered with yes, the result is classified as non-public. Business or trade secrets need to be removed in accordance with all partners' requirements before it can be published.

6. **Does a result name technologies that are part of an ongoing, project-related patent application?**

If this question is answered with yes, then the result is classified as non-public. Of course, results can be published after the patent has been filed.

7. **Does a result break the security interests of any project partner?**

If this question is answered with yes, the result is classified as non-public.

This is a simple structural approach to determine the different data types defined as part of the DMP. The responsibilities of the FORTESIE consortium partners towards disseminating the project outcomes are defined in the following section.

2.3.2 Data Archiving

As the FORTESIE technical solution gets more mature, more and more data will be ingested to the platform. Such data can be public-interest data (e.g., weather data), operational data from smart meters in pilot sites, or personal data from new users. For the first two categories, data will be ingested to the systems at very high frequencies, resulting in large volumes of data. On the other hand, such data tend to be very useful when they are fresh for the development of real-time services, while after a short time from their ingestion they are mostly used for static analysis and for batch processing analytics. Moreover, the probability to update such data is very low, especially after some days. Such data will be available (only read permissions) in batches to users only if they are authorised to have access to it. Access to data will be granted via an identity management component. This component makes sure that a user is authenticated to FORTESIE and provides access to the requested resources only if access policies to the requested resources are in line with the request. Of course, access policies for each dataset will be decided by the data owners.

Regarding personal data, they will be stored for a period and then removed. The personal data collected as part of the project will be limited to the project submission and informed consent of participants about the use of personal data will be required. Personal identity will be protected by the use of anonymous codes. The relation of real names and codes will only be known to project partners who will keep the records in a secure place. Later in the project's duration, the consortium will discuss the possibility to ensure that some data after the project will be anonymised and become available for further research purposes. As such, FORTESIE will generate two distinct consent forms: one for the data gathering activities of the pilot demos and one for after the project's duration. The purpose of the latter consent form will be to receive the data owners' consent to anonymise the data that they provide and make them available for further research purposes (which will be detailed explicitly in the consent form).

2.3.3 Data Performance

As already mentioned in Section 2.3.2, FORTESIE will make use of large amounts of different types of data coming from heterogeneous sources and providers. Amongst others, the continuous availability of such data is an indisputable attribute that entails high computational and performance requirements.

The FORTESIE complete framework will be deployed at pilot sites through controlled environments, decoupled from 'production' environments, with the use of a dedicated data processing infrastructure for experimental purposes exploiting large volumes of historic and live data, possibly anonymised, or simulated. This will alleviate the burden of using only the computational resources of an integrated system and will transfer the computational load of data processing, analysis, etc. to the pilot sites.

In addition, the FORTESIE platform envisages an architecture that separately addresses each stage of the data flow within the platform i.e., data interoperability and homogenisation, data streaming, and data storage.

FORTESIE will gather an abundance of data from IoT and other sensors, historical data, data from pilots' private databases, open data from public databases, and data from questionnaires. Among others. Regarding data interoperability and homogenisation of IOT and sensor data, data captured will be translated by IoT and System adapters to NGSI (Next Generation Service Interfaces) data models and stored in the FIWARE Data Broker[12]. Data Homogenization will be carried on in this layer

[12] https://www.fiware.org/about-us/

and, to the extent that it is possible, community smart data models will be used. In this way, the upper layers of this architecture, where the processing is held, consist of existing software tools or newly developed ones that will be reusable by the community, ensuring standardisation in the industry. The upper layer will consist of monitoring and simulation components, task scheduling, reasoning, analytics, and optimization engines, where each service can be granted access to specific data that is relevant to its purpose and can transfer back processed data to the unified data collection system. Data collected to the data broker from all layers will be subject to data sovereignty rules. Permissions to access data will be individually granted or revoked to every data consumer. Authorization components will be used that comply with industry security standards.

2.3.4 Data Protection and Security

Processing of personal data will follow the REGULATION (EU) 2016/679 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and until valid, the repealing Directive 95/46/EC (General Data Protection Regulation - GDPR).

Based on the outcomes of the Data Protection Impact Assessment (DPIA), as provided for in Article 35 (1) of the GDPR, appropriate organisational and technical measures will be implemented to safeguard the protection of personal data (i.e. anonymisation, pseudonymisation, encryption at rest and in transit, hashing, tokenization, key management practices that protect data across all applications and platforms, etc). The personal data collected as part of the project will be limited to the project scope and informed consent from participants, about the use of personal data, will be required.

Internal operational data collected or generated from/by partners throughout the whole period of the project will be held in data repositories in the respective servers of each partner. The servers will be kept in locked rooms with strict access mechanisms adhering to appropriate security standards while making use of state-of-the-art security mechanisms. Access to the repositories will be allowed only to authorized personnel both at the physical and network level. Where datasets are stored in databases, access will be allowed only to authorized users, provided a unique username and password, following access privilege rules. Backups of the databases will be stored encrypted and on the premises of each respective organization. Data backups of devices will happen regularly and will be stored in devices that will follow the same security standards and procedures as the main server.

Transfer of data will follow established good practices, such as encrypting files and sharing the keys/passwords via secure means.

Processing of document-based data will be supported via a dedicated platform (ProofHub). According to their privacy policy[13] and security information[14]:

- it employs "*state of the art technology to maintain high standards of data security and ensure that … communications are secure, and businesses are protected*",
- "*all data is encrypted via SSL/TLS when transmitted*",
- "*On hourly basis, data gets backed up and copies of the data are saved and secured at an off-site location for disaster recovery"*[15], *while* "*database backups are encrypted*",
- "*items and files deleted are moved to trash from where they are purged after 15 days*" unless we empty the trash manually in which case "*data is purged immediately*",

[13] https://www.proofhub.com/privacy
[14] https://www.proofhub.com/security
[15] https://help.proofhub.com/plus/account/security-backup-data/

- *"data is saved on reliable servers and written to multiple disks and stored in multiple places to remove even the minutest point of failure"*,
- access to data is granted "*only to authorized team members*",

Additionally, ProofHub allows document-based data to be available in a read-only or downloadable format, hindering access to information by unauthorised users. Moreover, it supports file versioning[16].

Documents with restricted access will remain in a locked cabinet at the organisation's premises.

2.3.5 FAIR Data

The international FAIR Principles have been formulated as a set of guidelines for the reuse of research data. The acronym FAIR stands for findable, accessible, interoperable, and reusable research data.

FORTESIE takes the opportunity to contribute towards the acceleration of materialization of GAIA-X[17] in the energy domain and therefore, its architecture supports effective and trusted sharing of data among participants covering all requirements to support future data marketplaces:

(A) Data Interoperability: open source, standardized, and domain agnostic NGSI API ensures the interoperability of data between different systems.

(B) Data Sovereignty and Trust: The Identity Management of FIWARE allows identification, authentication, and authorization of organizations, and individuals, while IDS Connector facilitates trusted data exchange.

(C) Data value creation: FIWARE NGSI Marketplace will be used to: (i) define new data asset types; (ii) register offerings which typically means providing the description of the asset, the data models, the endpoints, the terms, and conditions of exchanging data including SLAs, legal clauses, and pricing schema; (iii) ability to navigate and search/discover existing offerings based on selected criteria.

All the aforementioned FIWARE functionalities ensure data manipulation in FORTESIE based on FAIR principles (Findability, Accessibility, Interoperability, and Reusability).

2.3.5.1 Making data findable

Storage, processing and sharing (among project participants) will be supported via a dedicated platform (ProofHub), whereas interaction with the wider public will be achieved through the official project website. Also, data will be stored at the coordinator's private cloud infrastructure repository and will be kept for a minimum of 5 years after the end of the project. Where requested, data will be kept for 2 more years.

A naming convention will include a concise description of contents, the host institution collecting the data and the month of publication.

Version numbering will only be an issue if a participant requests withdrawal of their data in which case a version number will be added to the filename.

Appropriate technical measures will be implemented ensuring that data will not identify any individuals and therefore real names of participants will not be distributed.

FORTESIE exploits building performance data in relation to some user actions, data collected (or metadata created) by different sources (mostly collected by building monitoring sensors and

[16] https://help.proofhub.com/plus/files/file-versioning-2/
[17] https://www.data-infrastructure.eu/GAIAX/Navigation/EN/Home/home.html

actuators and user interactions by a mobile app) during the period of renovation/digitisation interventions.

FORTESIE's data sovereignty module will adopt and extend FIWARE components for data manipulation and sharing to create a data space defining open data models, standard APIs, and viable data sharing policies (FAIR), and the Orion context broker[18] will be adopted and extended with services for data IoT data devices (gateways) management and remote control (for actuators). Throughout the FORTESIE project, FIWARE NGSI Marketplace will be used to: (i) define new data asset types; (ii) register offerings which typically means providing the description of the asset, the data models, the endpoints, the terms and conditions of exchanging data including SLAs, legal clauses, and pricing schema; (iii) ability to navigate and search/discover existing offerings based on selected criteria.

2.3.5.1.1 Discoverability of the data

Taking into account the FAIR data principles (meta)data will:

- Be assigned to a globally unique and persistent identifier;
- Contain enough metadata to fully interpret the data, and;
- Be indexed in a searchable source.

2.3.5.1.2 Data identification mechanisms

All documents will be identified by the project name, followed by a unique and persistent document type designator and number provided by the coordinator for the submission to the European Commission (EC). Versioning of the document should be part of the document name and title.

Document identification will include the task or deliverable number, used to identify the document, followed by a brief title of the activity or deliverable.

FORTESIE will also publish data through scientific articles, in which case DOIs will be provided from the publisher. For other literature, such as reports and policy recommendation, DOIs will be assigned via the repository in which they will be archived (e.g., Zenodo).

2.3.5.1.3 Naming conventions

To **(i)** enhance data searchability and discoverability, and **(ii)** provide clues to the content, status, and versioning of the files, each set of data produced (dataset, deliverables, etc…) will be named in a uniform way and will include a table with a version control.

The recommendations to name the documents of the project are as follows:

- Choose easily readable identifier names (short and meaningful);
- Do not use acronyms that are not widely accepted;
- Do not use abbreviations or contractions;
- Avoid language-specific or non-alphanumeric characters;
- Add a two-digit numeric suffix to identify new versions of one document.
- Dates should be included back to front and include the four-digit years: YYYYMMDD.

[18] https://fiware-orion.readthedocs.io/en/master/

For deliverables: **Project's name - Dx.y - [Name of the deliverable as described in the DoA]** being x - work package assigned to the deliverable y - the number of deliverables within the work package i.e.: D.1.2 - Data management plan. Report and updates M6.

For datasets: **Project's name - WP [Work Package number] P [Pilot number; pilot activity number] - T [Task number; description of the activity]** e.g., WP3 Task 3.3 Integration and technical testing.

Easy-to-use search keywords will be used in FORTESIE to optimise the reuse of data by interested stakeholders. The metadata standards employed by FORTESIE provide opportunities for tagging the data collected/generated and its content with keywords.

In general, the keywords will comprise terms related to the topics addressed, such as energy efficiency, energy renovations, smart contracting, innovative business models, fair energy transition, capacity building in the energy sector, green currency, smart renovations, energy efficiency policies, as well as keywords specific to the project, such as FORTESIE, Horizon Europe, etc.

The keywords will accurately reflect the content of the datasets and avoid words used only once or twice within them.

2.3.5.2 Making data openly accessible

Data will be made available where possible, subject to ethics and participant agreement. FORTESIE will use FIWARE Identity Management via the data sovereignty module, which allows identification, authentication, and authorization of organizations and individuals, while IDS Connector[19] facilitates trusted data exchange. As a result, data will be both accessible and trustworthy.

2.3.5.3 Making data interoperable

The concept of interoperability necessitates machine-readable data and the use of consistent terminology. FORTESIE supports the Data Interoperability principle by providing an open source, standardized, and domain agnostic NGSI API that ensures data interoperability across systems.

2.3.5.4 Increase data re-use

2.3.5.4.1 Increase data re-use through clarifying licences

The use of Creative Commons licences, the default being CC-BY, will ensure that data will be widely re-usable. This licence is used for research articles, allowing copying, distribution and transmission of work without affecting key author rights. The re-use of data (if needed) will be restricted to the research use of the license and anonymous data can be used for scientific publications. Data may not be copied or distributed and must be referenced if used in publications. The collected data will be a consolidation of data from several sources, each one having its own policies.

2.3.5.4.2 Data quality assurance process

The data quality principle comprises that data has to be of good quality, i.e., the data has to be accurate and up to date. This implies that personal data processing will be done following the EU, national and international laws taking into account the "data quality" principles listed below:

- Data processing is adequate, relevant and non-excessive.

[19] https://internationaldataspaces.org/offers/ids-components/

- Accurate and kept up to date.
- Processed fairly and lawfully.
- Processed in line with data subjects' rights.
- Processed in a secure manner.
- Kept for no longer that necessary and for the sole purpose of the project.

The data quality assurance process will be led in accordance with the REGULATION (EU) 2016/679 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data.

2.3.5.4.3 Length of the time in which the data will remain re-usable

The Consortium will contribute to maintaining data reusable as long as possible after the end of the project. A first period of 5 years has been established; however, this time can be extended under a partner's agreement. This period can vary depending on the value of the data after the end of the project.

2.3.6 Allocation of resources

Publication of the FORTESIE results and assets to the aforementioned publishing platforms in a way that makes them FAIR in most cases will not require extra costs as these services are provided for free to its users.

2.3.7 Privacy and Data Protection

2.3.7.1 Removing Personal Identifiers

Datasets will be anonymised for impact assessment and research purposes. The personal data collected as part of the project will be limited to the project submission and informed consent of participants about the use of personal data will be required. Personal identity will be protected using anonymous codes. The relation of real names and codes will only be known to project partners who will keep the records in a secure place. The relation of applications will be coded and will be available for external evaluators with such coding. In case data needs to be transferred to non-EU partners, we will obtain approvals from the competent Data Protection Office, unless those countries are on the list of countries that provide adequate safeguards with respect to the protection of the privacy and fundamental rights and freedoms of individuals and as regards the exercise of the corresponding rights. All copies of approvals /notifications regarding the processing of personal data by the competent institutional Data Protection Office will be made available upon request to the EC. Personal data will be encrypted and stored securely. The personal data protection processes will follow the REGULATION (EU) 2016/679 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 27 April 2016 on the protection of natural persons regarding the processing of personal data and on the free movement of such data, and until valid, the repealing Directive 95/46/EC (General Data Protection Regulation).

2.4 Data Archiving and Preserving Infrastructure

Along with the definition of the datasets, special focus is delivered on the selection of the platform to archive and preserve the datasets. When we choose a repository, it is important to consider factors such as whether the repository[20]:

- Gives the submitted dataset a persistent and unique identifier. This is essential for sustainable citations – both for data and publications – and to make sure that research outputs in disparate repositories can be linked back to particular researchers and grants.
- Provides a landing page for each dataset, with metadata that helps others find it, tell what it is, relate it to publications, and cite it. This makes research more visible and stimulates the reuse of the data.
- Helps to track how the data has been used by providing access and download statistics.
- Responds to community needs and is preferably certified as a 'trustworthy data repository', with an explicit ambition to keep the data available in the long term.
- Matches particular data needs (e.g.: formats accepted; access, backup and recovery, and sustainability of the service). Most of this information should be contained within the data repository's policy pages.
- Provides guidance on how to cite the data that has been deposited.

2.4.1 FORTESIE Website

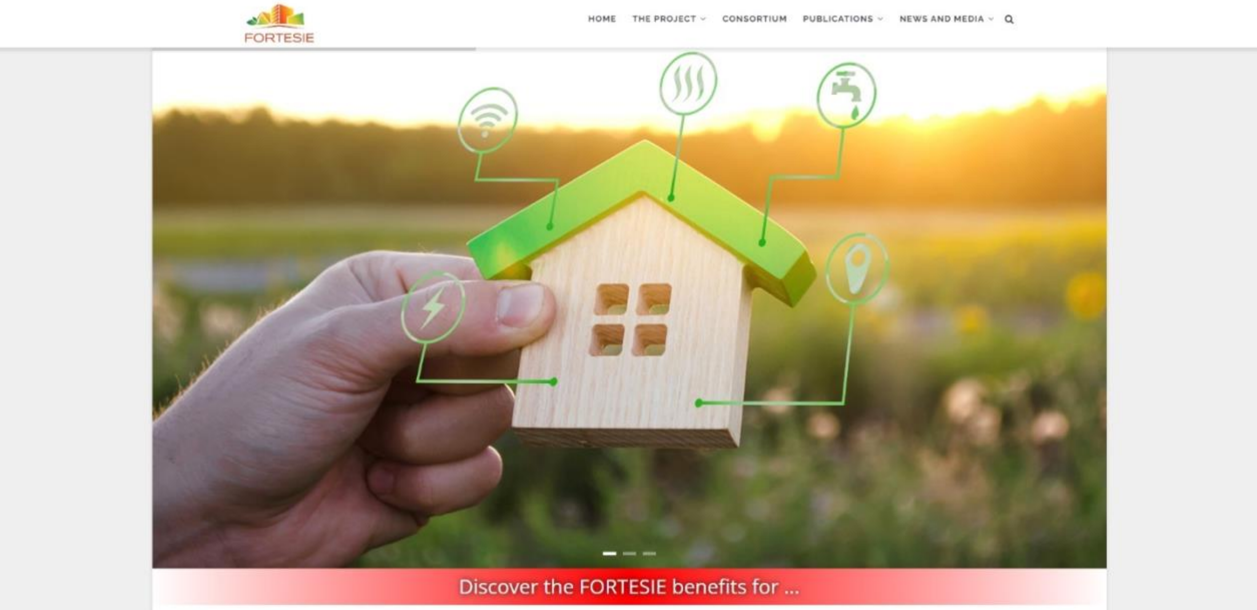
The FORTESIE consortium decided early to set up its own project-related webpage. This webpage describes the objectives and the general approach of the project, the partners, the pilots, and its development status. A "news and media" tab informs about news on a regular basis. A dedicated section for publications (giving the opportunity of downloading) is used to publish public deliverables, reports, and white papers.

All documents are published using the portable document format (PDF). All downloads will be enriched by using simple metadata information like the title and the type of the document. The webpage was designed and developed by the partner of the consortium INCL.

All webpage-related data is backed on a regular basis by INCL. All information on the FORTESIE website can be accessed without creating an account. The webpage will be backed up at regular intervals by INCL.

The FORTESIE webpage will be available during the project runtime and will still be available for at least two years after the official project end.

[20] How to select a data repository? https://www.openaire.eu/opendatapilot-repository-guide



Web link: http://fortesie.eu/

2.4.2 ProofHub

ProofHub[21] is a project planning software that includes several tools for team cooperation, a calendar for tasks and deadlines, file repositories, chatting functionalities and the possibility to create different topics for parallel streams of activities. It is a web-based browser application, developed by ProofHub LLC in 2011.

ED as the project coordinator has purchased ProofHub "as a service" and the FORTESIE project will use it to store project related data internally in the system. Access to the FORTESIE ProofHub is controlled by ED and given only to authenticated FORTESIE partners.

ProofHub Link: https://eurodyn.proofhub.com/bapplite/#app/overview/project-6585418367

2.4.3 Zenodo

Zenodo[22] is a research data archive/ online repository which helps researchers share research results in a wide variety of formats for all fields of science. It was created through EC's OpenAIRE+ project[23] and is now hosted at CERN using one of Europe's most reliable hardware infrastructures. Data are backed nightly and replicated to different locations. Zenodo supports not only the publication of scientific papers or white papers, but also the publication of any structured research data (e.g., using XML). Zenodo provides a connector to GitHub that supports open collaboration for source code and versioning for all kinds of data. All uploaded results are structured by using metadata, like for example the contributors' names, keywords, date, location, kind of document, license, and others. Considering the language of textual metadata items, English is preferred. All metadata is licensed under CC license

[21] https://www.proofhub.com/
[22] https://en.wikipedia.org/wiki/Zenodo
[23] https://www.openaire.eu/

(Creative Commons 'No Rights Reserved'[24]). The property rights or ownership of a result does not change by uploading it to Zenodo.

All public results related to scientific publications that will be produced during the FORTESIE project will be uploaded to Zenodo for long-term storage and open access.

Project Zenodo Link: Will be provided in the following update of the deliverable.

2.4.4 Code Repositories: GitHub

FORTESIE will use two different types of repositories for the programming code that it will generate under the context of the FORTESIE technical solution.

Private tools will be stored in one or more private repositories or infrastructure belonging to specific project partners, providing access to all consortium members or just the members to whom a specific tool belongs.

For open-source components, the FORTESIE technical team will explore various options of open code repositories such as GitHub.

GitHub[25] is a well-established online repository that supports distributed source code development, management, and revision control. It is primarily used for source code data. It enables worldwide collaboration between developers and provides some facilities to work on documentation and track issues.

GitHub provides paid and free service plans. Free service plans can have any number of public, Open Access repositories with unlimited collaborators. Private, non-public repositories require a paid service plan. Many open-source projects use GitHub to share their results for free. The platform uses metadata like contributors' nicknames, keywords, time, and data file types to structure the projects and their results. The terms of service state that no intellectual property rights are claimed by the GitHub Inc. over provided material. For textual metadata items, English is preferred.

Web link: https://github.com/

Project GitHub Link: Will be provided in a following update of the deliverable.

2.4.5 Project Communication Channels

Besides the FORTESIE website, project-specific Web 2.0 channels have been launched aiming at extending the visibility of the project's activity. These include FORTESIE accounts on:

- LinkedIn: https://www.linkedin.com/company/fortesie-horizoneu/
- Facebook: https://www.facebook.com/profile.php?id=100087107495674
- Instagram: https://www.instagram.com/fortesie_horizoneu/

2.4.6 FORTESIE Platform

From a data management perspective, FORTESIE will be a platform in which several datasets (structured or unstructured) from different energy data sources (e.g., sensors, IoT devices, smart meters, etc.) are ingested on a daily basis, either in batches (batch data ingestion) to facilitate aggregate analytics services, and services based on historical data, or through data streaming

[24] https://creativecommons.org/share-your-work/public-domain/cc0/
[25] https://en.wikipedia.org/wiki/GitHub

technologies, to facilitate near real-time services. As a next step, data will be processed, in order to improve their quality, and homogenised and modelled, in order to be efficiently shared with users or sent to the data analytics services in an understandable format. After this step, data are transferred to a storage, to be queried and utilised by energy analytics services and users.

Of course, a security and access control component, on top of the data management services is of paramount importance. This component is responsible for securing that only authenticated and authorised users and services can have access to the requested resources. So, if a user is not logged in to the platform, the access control component will prohibit access to the requested resources. The same applies to authenticated users that try to access a resource (data or service) and do not have permission to the requested resource. For this functionality user data will be stored in a relational database. Regarding security, the provided security framework will offer data encryption, vulnerability detection and mitigation, as well as user behaviour monitoring and auditing. No more specific details can be given at this point about the FORTESIE platform as it is still early in the project's duration and all related conversations are theoretical and preparatory.

2.5 Intellectual Property Rights (IPR) Guidelines in the context of FORTESIE

In the context of the FORTESIE project, appropriate IPR handling is very important to maximise exploitation and dissemination results and overall projects' outcomes while at the same time ensuring the protection of intellectual property. Therefore, it is crucial to identify the IPR of the datasets, the software, the tools, and the knowledge that will be used or produced in the project. This includes licensing schemes, terms of usage and access rights of these assets.

IPR handling is addressed privately within the project. In brief, it concerns IPR management at the Consortium level, covering issues relating to access to the background, ownership of foreground and results, access rights and IPR protection regarding dissemination activities.

In addition, activities of WP5 and more specifically Task 5.4: Exploitation Planning will provide details of IPR and ownership of results to safeguard the rights of all partners. In particular, Deliverable D5.4 and D5.5 will provide information on the ownership of background and foreground. In addition, it will describe how IPR will be safeguarded after the project's completion and concerns the ownership of the results that may be generated after the end of the grant period.

Furthermore, the obligations, rights and responsibilities of partners are described in Consortium Agreement (CA) signed by all partners. Issues regarding IPR, such as joint ownership of results or cases in which IPR are affected, are also prescribed in CA. In addition, details about the background that each partner brings into the project are described in CA.

3 FORTESIE Datasets

In this section, a list of all existing or foreseeable results for dissemination is presented, separated into public deliverables, Private and Open-Source software components, publications, and pilot-related data. For each result and in accordance with the FAIR data management guideline we provide a description and name of the standards used for storage and metadata (to make data findable & interoperable).

3.1 Dataset I: FORTESIE List of Public Deliverables

We are considering the FORTESIE Project public deliverables as part of the data management plan. The following table presents the list of public deliverables of the FORTESIE project.

Table 3-1: List of FORTESIE public deliverables

D_ID	Title	Due Date
1.1	Quality, risk and innovation handbook	M3
1.2	Data management plan. Report and updates	M6
1.4	Data management plan. Report and updates M36	M36
2.1	End-user and pilot requirements and use-cases description	M6
2.2	FORTESIE services co-creation M9	M9
2.3	Reference architecture and components functionality M9	M9
2.4	Business models analysis for each service M9	M9
2.5	FORTESIE services co-creation M25	M25
2.6	Reference architecture and components functionality M25	M25
2.7	Business models analysis for each service M25	M25
3.4,5	Report on the renovation technologies selected for each building, and deployment approach	M20, M25
4.2	Engagement Plan and Social Acceptance assessment	M18
4.3,4	Pilots execution documentation and validation assessment	M28, M34
5.3	Interaction with European projects and strategies, ECB and DeFi communities	M36
5.4,5	Online OSS demonstrated and exploitation planning and IPR report including alternative financing schemes	M18, M36
6.1	Communication and Dissemination Plan	M3

6.2	Website operational and promotional materials	M3
6.3	Website and Communication, dissemination, and stakeholder engagement report M18	M18
6.4	White Paper: lessons learnt and policy recommendations	M36
6.5	Website, Communication, dissemination and stakeholder engagement report M36	M36

3.2 Dataset II: EU Classified Deliverables: Handling of EU Classified Information (EUCI)

The FORTESIE Consortium is obliged to respect the security rules for protecting EU Classified Information (EUCI), as laid down in Commission Decision (EU, Euratom) 2015/444. This means that the FORTESIE partners will handle EUCI in line with the principles, rules and procedures established therein. By compromising or losing EUCI we are aware that we do not only breach the grant agreement, but we are also liable to disciplinary and/or legal action in accordance with applicable (national) laws, rules and regulations.

For information marked as such no Personnel Security Clearance Certificates need to be issued according to the aforementioned Euratom Decision.

However, the FORTESIE Consortium is aware that EU RESTRICTED deliverables require special treatment:

- **Need-to-know**: To access EUCI, individuals need to have a need-to-know (i.e., you need the information to perform a specific professional function or task).
- **Awareness of the Security Rules**: Individuals may only be granted access after they have been briefed on the security rules and have acknowledged their responsibilities.
- **By post**: Documents should be sent in double, opaque envelopes. The inner envelope must be sealed and marked RESTREINT UE/EU RESTRICTED. The documents/USB key/CD ROM inside must bear the same marking.
- **Electronic transmission**: The documents are encrypted with approved encryption tools (i.e., FILKRYPTO) and sent via e-mail. EU classified deliverables shall not be uploaded via the portal.
- **Storage**: When not in use, the documents shall be stored in a locked container or a locked cupboard.
- **Consultation**: Only in secure areas where nobody is able to read or remove the documents.
- **Printing/copying**: Printing and/or copying of documents is not allowed.
- **Notes**: Reference to EUCI in the partners' notes means that the notes should be treated as RESTREINT UE/EU RESTRICTED documents.
- **Communication**: There shall be no communication about the EUCI via phone, e-mail or in unsecured areas. The information shall not be discussed with persons who do not have a justified need-to-know.
- **Meetings**: Meetings on EUCI are organised on an invitation-only basis and all attendees need to have a need-to-know. Attendees should sign an attendance sheet at the beginning of a

meeting. Meetings should take place behind closed doors. The windows and blinds in the room must be closed. Mobile phones and other portable devices must be switched off or left outside of the meeting room. Wireless microphones may not be used, and microphones/speakers should be switched off. If RESTREINT UE/EU RESTRICTED documents are distributed at the beginning of a meeting, the exact number of documents necessary must be distributed and they must be returned to the Chair, collected, and accounted for. No stock may be held in the room. If EUCI is presented or projected on a screen, the computer or laptop used must not be connected to a network. During breaks, meeting rooms must be locked and guarded. Feedback on EUCI must be raised during the specific time slot dedicated to its discussion (on the agenda).

- **Breach or compromise**: In case of a (potential) breach or compromise of EUCI, the partner/partners must immediately inform the REA SPOC TEAM (REA EUCI SPOC REA-EUCI-SPOC@ec.europa.eu) and the responsible REA Project Officer. No attachment of EU Classified Information shall occur during this communication.
- **Duration**: Obligations vis-à-vis the protection of EU classified information continue to persist after the life of the project.

3.3 Dataset III: Private and Open-source Software Components

The analysis is performed by considering the list of exploitable components. An indicative list, to be updated (with additions, deletions, etc.) after architecture finalisation, is presented in the following table. Additionally, sections 3.3.1-3.37 present some early, indicative information regarding the inputs, outputs, and hyperparameters of each component.

Table 3-2: FORTESIE Software Components

Title	Responsible Partner
Data Sovereignty Module	ED
Gamified Mobile App	ED
Green Euro-Neobanking App	CCO2
Behavioural model and recommendation engine	SIN, NTUA
Blockchain/Smart Contract based M&V calculation module	CTIC, CCO2
Energy Performance Assessment and Certification Service	CTIC
Online renovation & ESIE Marketplace	ED

3.3.1 Data Sovereignty Module

Table 3-3: Data Sovereignty Module Data Requirements

Data Requirements	
Input data	Data from all sensing devices, users input (i.e., Users' profiles, home data, number of residents, devices owned, and data for personalized recommendations) and external systems (i.e., local weather)
Output data	All input data transformed to the commonly agreed model and compliant with NGSI agents. They will be provided to all modules which process data, including almost all of the above, except the online renovation & ESIE marketplace, which will deal with separate data.
Hyperparameters	• Stored and made available for all the processing needs of identified/selected system components. • Managed with the FAIR principle

3.3.2 Gamified Mobile Application

Table 3-4: Gamified Mobile Application Data Requirements

Data Requirements	
Input data	a. Users' input (i.e., registration of new devices, inserting necessary information where needed, profile and building description, performance input (actions, or feedback on recommendations, etc.)) b. Data/measurements from the meters/sensors installed in the residence. c. Newly processed data produced by the FORTESIE platform, using data coming from a. and b. above
Output data	Processed data. • Visualization data (i.e., energy consumption for specific time intervals, Savings Display, CO_2 available/gained/spent coins, etc.) • Personalised recommendations, game challenges and rewards gained. • Help with complex tasks and tips for improving performance.
Hyperparameters	• **Monitor the ESIE performance** (daily, weekly, monthly) for the identified measurements for each pilot. • Check **graphs** that analyse the ESIE performance compared to a **baseline** (daily, weekly, monthly) • Check the individual sensors' measurements e.g., **indoor and outdoor** (regional) **temperature, humidity, energy consumption, energy bill, etc.** • Receive **personalised messages and recommendations** related to ESIE performance improvements. • **Earn Green Euros badges** of different categories (streak, tree, monthly, quiz and profile) and "grow" the home page tree. • Check the earned Green Euros and Green Euros balance at any time. • Feedback on achieving ESIE: Easy to understand visualisation of ESIE performance data and achievements in graphics and Tamagochi figure (tree growth) • Personalised challenges to engage users in actions (vote for reducing comfort levels, or behavioural change actions) • Comparison of challenges to measured achievements. • Allocation of batches and interaction with the game

3.3.3 Green-Euro Neobanking Application

Table 3-5: Green Euro Neobanking Application Data Requirements

Data Requirements	
Input data	a. KYC (Know Your Customer) - a legal and mandatory process, part of European regulation. Includes documents and information from your end users (legal information, proof of identity, proof of residency, delivery point, etc.) b. Consumptions Data from partners (type of heating systems, real consumptions, PDL, etc.) c. Proof of consumption (bills, other certified documents)
Output data	• A banking account & credit card (with 0 charges in the euro zone for card payments or transfers) where GE earned are automatically made available. • A banking application • Dashboard on CO_2 emission performance including timelines. • A simple application (open source and open data) to estimate, compute and compare carbon footprints (for housing, transport, and food habits) • Personalised recommendations to engage users in actions (renovations, heating system improvements, transports). • Personalised challenges leading to €G rewards in order to reinforce user engagement. • A personalised API to each of our partners to proceed with agreed actions automatically & consult data (e.g., earned, badges, etc.).
Hyperparameters	• CO_2 credit transactions and monitoring. • Study a €G smart contract implementation. • A loan offer around 0.5% APR for projects to reduce users' CO_2 emissions.

3.3.4 Behavioural Model and Recommendation Engine

Table 3-6: Behavioural Model and Recommendation Engine Data Requirements

Data Requirements	
Input data	a. Users' input (i.e., registration of new devices, inserting necessary information where needed, profile and building description, performance input (actions, or feedback on recommendations), preferences, actions, stakeholders' reactions to proposed energy solutions, etc.). b. Data/measurements from the meters/sensors installed in the residence. c. Data measurements from other mobile devices (e.g., smartphones, smartwatches, etc.) to prove the behaviour/mobility of a user. d. Newly processed data produced by the FORTESIE platform, using data coming from a., b., and c. above.
Output data	Engagement recommendations and personalised interaction flows which create user feedback loops that trigger meaningful changes in user behaviour, based on tailored UX, content design, and incentivization. These will indicatively include: • Education material that will empower participants to learn more about how they can optimize their energy consumption behaviour, thus encouraging ongoing participation and continued app use. • Social rewards for participants, as people want to show that they care and that they behaved according to virtuous values. • Motivation pathways for energy savings.

	• Enlargement of results by nudging participants towards a bigger, common goal creating a community approach. • AI-based information, aiming at finer-grained energy profiling & forecasting, energy resources management, enhanced comfort & people's well-being. • Personalised suggestions, taking into account individual energy consumption, energy cost, carbon emissions, renewables production and thermal comfort.
Hyperparameters	• **Monitor the user performance** (daily, weekly, monthly) for the identified incentives for each user. • Check **graphs** that analyse the user performance compared to a **baseline** (daily, weekly, monthly). • Check the individual sensors' measurements e.g., **indoor and outdoor** (regional) **temperature, humidity, energy consumption, energy bill, etc.** to ensure that users are sticking to the proposed plan. • Feedback on achieving ESIE and different pathways based on the personal criteria of users to achieve the necessary savings. • Challenges to engaging users to behavioural change actions.

3.3.5 Blockchain/Smart Contract based M&V Calculation Module

Table 3-7: Blockchain/Smart Contract based M&V Calculation Module Data Requirements

Data Requirements	
Input data	a. Measurements of EPC sensors b. Smart performance contracts
Output data	Verification of EPC sensors' measurements (before and after)
Hyperparameters	To be decided at a later stage

3.3.6 Energy Performance Assessment and Certification Service

Table 3-8: Energy Performance Assessment and Certification Service Data Requirements

Data Requirements	
Input data	Real energy consumption data, distinguishing among the consumption derived from the energy fabric, energy systems and user behaviour.
Output data	Comparative assessment of the performance of buildings from an energy and sustainability perspective.
Hyperparameters	To be decided at a later stage

3.3.7 Online Renovation & ESIE Marketplace

- Dataset description (input/output)
- How the data will be used (service operation)

Table 3-9: Online Renovation & ESIE Marketplace Data Requirements

Data Requirements	
Input data	a. Data provided by the users to inform about ESIE policies, available actions, and evidence collected per target group, and recommend active players in each region to contact for further support.

	b. Financing mechanisms and programmes available in each country. c. Success stories.
Output data	Feedback on visits, successful customers, contracts, and new projects initiated.
Hyperparameters	To be decided at a later stage.

3.4 FORTESIE Scientific Publications

Along with the dissemination of project deliverables and datasets, we are considering as part of the Data Management Plan, further dissemination of project Scientific Publications. As of now, there have been no scientific publications generated under the context of FORTESIE.

3.5 FORTESIE Pilots Data

The following data will be collected from the project's pilots, based on a common FAIR-related questionnaire. They present the pilots' first approach to FAIR principles and are going to be updated, according to the project's progress. The following table provides a central approach for pilot data gathering, given that the project is aiming to develop a central/horizontal technical solution. Many of the rows of the table are not possible to be completed at this early stage of the project.

Table 3-10: FORTESIE Data Collection Template

FORTESIE Data Collection **Data Outline**	
What is the origin of the data?	Energy Meters, Smart Meters, Environmental Sensors, Controllers (VFDs), Electrical bills
What are the purpose and utility/usefulness of the data collection/generation?	Statistics, Energy Profile Creation, Energy Consumption Monitoring and Optimization Logic for AHU Fun speed control, Improving current EPC
What is the direct or indirect relation to the objectives of the project?	Direct relation to reducing energy consumption and CO_2 footprint; provide real results and values to be able to improve the IPMVP practices
What are the types and formats of the data collected/generated?	csv, xlsx, JSON
What is the expected/estimated size of the data (approximately, if known)?	To be determined at a later stage.
FAIR Data ***Making data findable, including provisions for metadata***	
What metadata will you provide to support the discoverability of data?	To be determined at a later stage.
How data will be identifiable? What standard identification mechanisms will be used (if any)?	To be determined at a later stage.

Will you make use of persistent and unique identifiers, such as Digital Object Identifiers?	
What naming conventions will be used?	The nomenclature of the data will be the sum of the pilot's name, the year, and the month to which the file refers. Pilot’s Name-Year-Month.csv
What approach for clear versioning will be followed?	To be determined at a later stage.
What standards for metadata creation will be used (if any)? Otherwise, what type of metadata will be created and how?	To be determined at a later stage.
FAIR Data ***Making data openly accessible***	
What data will be made openly available? What is the rationale behind data being kept closed (if any)?	Energy consumption data, indoor environmental air quality data
How the data will be made available?	To be determined at a later stage (indicatively, information could be shared via shared folders on the cloud or the project’s data repository..
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 Sourcecode)?	To be determined at a later stage.
Where the data and associated metadata, documentation and code are deposited? Are there any privacy issues associated with this process?	To be determined at a later stage.
How access will be provided in case there are any restrictions?	In case of restrictions, the pilot leader will directly share/provide the data.
FAIR Data ***Making data interoperable***	
What data and metadata vocabularies, standards or methodologies you will follow to facilitate interoperability?	To be determined at a later stage.

Will you be using standard vocabulary (or more commonly used ontologies) for all data types present in your data set, to allow interoperability?	To be determined at a later stage.
FAIR Data ***Data Security***	
How will you address data recovery, secure storage, and transfer of sensitive data?	To be determined at a later stage.

4 Data Management Implementation

This section provides an overview of the specific provisions, foreseen for the management of each project dataset in table form.

Table 4-1: FORTESIE Data Management Implementation

FORTESIE Dataset	Classification	Archiving	Performance	Safety & Security	FAIR	Privacy & Data protection
FORTESIE Project Public deliverables	Public	Each public deliverable will be published openly on the FORTESIE webpage (following European Commission review and approval). All earlier versions of it will be archived on the project's internal ProofHub repository.	N/A	N/A	Public deliverables uploaded on the project website with the appropriate metadata	N/A
Open-Source Software Components	Public	Software developers will share their code base on their own public repository e.g., GitHub.	N/A	N/A	Source code deposited in GitHub or some other type of public repository	N/A
FORTESIE Scientific Publications	Public	FORTESIE Website, Zenodo	N/A	N/A	Publications indexed in the project website and the project's dedicated Zenodo with the appropriate metadata	N/A
FORTESIE Pilots' Public Data	Public	Data will be stored in the project DB in two different ways: i) near-real time for data coming from active sensors connected to the platform, and ii) in asynchronous mode for data coming from	N/A	Project DB will be under a private network accessible only to project members and secured under the cloud provider security policy.	Data will be translated to NGSI models and stored in the FIWARE Data Broker.	Potential anonymisation (suppression, generalisation, etc).

		different pilots' sources such as files or historical DBs.				
FORTESIE Pilots' Private Data	Confidential	Data will be stored in each pilot partner database (on premises)		Individual security mechanisms and policies of each beneficiary.	N/A	N/A

5 Conclusions

This document introduces the plan that the FORTESIE project will take for data management and provides an initial analysis of the data sources that will be used or generated during the project as identified by the project consortium partners and the way the project results will be shared. By project results this deliverable defines any kind of information including scientific publications, white papers, Open-Source code, open datasets, anonymous interview results, or mock-up datasets used for gathering customer feedback that may be used or generated from the project. The collected datasets in the current version of the report are research data, related to the project's work packages and are managed according to their level of availability (public or Consortium). The FORTESIE Data Management also follows the Guidelines on FAIR Data Management in Horizon 2020, i.e., data must be findable, accessible, interoperable, and reusable.

The current report will be a living document throughout the project. The DMP will be updated whenever significant changes arise, such as (but not limited to) new data, new innovations, patent filings, changes in the consortium members and others.

Appendix A: FAIR Data Management at a glance: issues to cover in your Horizon Europe DMP

This table provides a summary of the DMP issues to be addressed, as outlined above.

DMP component	Issues to be addressed
1. Data summary	• State the purpose of the data collection/generation. • Explain the relation to the objectives of the project. • Specify the types and formats of data generated/collected. • Specify if existing data is being re-used (if any). • Specify the origin of the data. • State the expected size of the data (if known). • Outline the data utility: to whom will it be useful?
2. FAIR Data 2.1. Making data findable, including provisions for metadata	• Outline the discoverability of data (metadata provision). • Outline the identifiability of data and refer to the standard identification mechanism. Do you make use of persistent and unique identifiers such as Digital Object Identifiers? • Outline naming conventions used. • Outline the approach towards search keywords. • Outline the approach for clear versioning. • Specify standards for metadata creation (if any). If there are no standards in your discipline describe what type of metadata will be created and how.
2.2 Making data openly accessible	• Specify which data will be made openly available. If some data is kept closed provide a rationale for doing so. • Specify how the data will be made available. • Specify 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)? • Specify where the data and associated metadata, documentation and code are deposited. • Specify how access will be provided in case there are any restrictions.
2.3. Making data interoperable	• Assess the interoperability of your data. Specify what data and metadata vocabularies, standards, or methodologies you will follow to facilitate interoperability. • Specify whether you will be using standard vocabulary for all data types present in your data set, to allow inter-disciplinary interoperability. If not, will you provide mapping to more commonly used ontologies?
2.4. Increase data re-use (through clarifying licences)	• Specify how the data will be licenced to permit the widest reuse possible. • Specify when the data will be made available for re-use. If applicable, specify why and for what period a data embargo is needed. • 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 reuse of some data is restricted, explain why. • Describe data quality assurance processes. • Specify the length of time for which the data will remain re-usable.
3. Allocation of resources	• Estimate the costs for making your data FAIR. Describe how you intend to cover these costs. • Clearly identify responsibilities for data management in your project. • Describe the costs and potential value of long-term preservation.
4. Data security	• Address data recovery as well as secure storage and transfer of sensitive data.

5. Ethical aspects	• To be covered in the context of the ethics review, ethics section of DoA and ethics deliverables. Include references and related technical aspects if not covered by the former.
6. Other	• Refer to other national/funder/sectorial/departmental procedures for data management that you are using (if any).