WeForming

DELIVERABLE 1.2

DATA MANAGEMENT PLAN

APRIL 8TH, 2024

Grant Agreement No.: 101123556

Topic: HORIZON-CL5-2022-D4-02
Type of action: IA

DOCUMENT INFORMATION

Project Acronym	WeForming
Project Full Title	Buildings as Efficient Interoperable Formers of Clean Energy Ecosystems
Work package	WP1 Project Management and Administration
Deliverable	D1.2 – Data Management plan
Due date	M6 - 31/03/2024
Document type	Report
Dissemination level	Public
Authors	ED
Reviewers	HSE, F6S

DOCUMENT REVISION HISTORY

Version	Date	Description of change	Author
V0.1	20.02.2024	Table of contents	ED
V0.2	20.03.2024	Full draft	ED
V0.3	29.03.2024	Internal review	ED
V0.4	03.04.2024	Final improvements	ED
V0.5	05.04.2024	First partner review	F6S
V0.6	05.04.2024	First review comments addressed	ED
V0.7	05.04.2024	Second partner review	HSE
V0.8	06.04.2024	Implementations of suggestions	ED
V0.9	06.04.2024	Quality review and editing	ED
V1.0	08.04.2024	Final version	ED

CONSORTIUM PARTNERS

LEGAL NOTICE

How to cite this report: WeForming (2024). D1.2. Data Management Plan.

Funded by the European Union

The WeForming project has received funding from the European Union's Horizon Europe Programme under the Grant Agreement No. 101123556.

UK Research and Innovation

The UK participant is co-funded by UKRI.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABBREVIATIONS

API	Application Programming Interface
BUC	Business Use Case
CIM	Context Information Management
DMP	Data Management Plan
EC	European Commission
ED	European Dynamics (Advanced Information Technology and Telecommunication Systems SA)
ETSI	European Telecommunications Standards Institute
EU	European Union
DoA	Description of Action
DOI(s)	Digital Object Identifier(s)
FAIR	Findable, Accessible, Interoperable and Reusable
GDPR	General Data Protection Regulation
iGFBs	Intelligent Grid-Forming Buildings
IPR	Intellectual Property Rights
ISG	Industry Specification Group
NGSI-LD	Next Generation Service Interface - Linked Data
SLA(s)	Service Level Agreement(s)
WP(s)	Work Package(s)

Executive Summary

This deliverable summarises the Data management and IPR protection procedures that will be followed during the course of the project. It provides the analysis of the Data management policy and the Data management lifecycle for the datasets that will be collected, processed or generated by the project. The plan is aligned with the guiding principles outlined in the Horizon Europe Framework Programme, particularly emphasizing the essential attributes of findable, accessible, interoperable, and reusable data. It incorporates these concepts within the framework of WeForming's objectives and the specific requirements of the project.

Central to this DMP is the life-cycle management of data generated and used within the project. This includes systematic processes for data identification, collection, processing, storage, utilization, sharing, archiving, and eventual disposal. The structure and workflow of the project provide the basis for identifying key data types and sources pertinent to the project's focus on energy-efficient and interoperable solutions for Intelligent Grid-Forming Efficient buildings (iGFBs).

In addition to the conventional data management life-cycle, this DMP addresses the unique aspects of the data requirements, including the handling of information pertinent to building operational optimization, energy management systems, and multi-sector interoperability. The plan contemplates the creation and management of data entries that align with the project's diverse information sources and stakeholders. These entries encompass a broad spectrum of information, ranging from digital operational data to interactions with the energy ecosystem.

To effectively manage this data, the DMP introduces a structured management process comprising initialization, assessment, management, and reporting phases. This iterative process is designed to evolve with the project, ensuring that data management practices remain effective and efficient. A key component of this process is a simplified data management template, facilitating consistent documentation of data sets and their life-cycle within the project.

For the practical implementation of the DMP, an issue tracking tool is proposed, which will assist in maintaining, reporting, and sharing information about the project's datasets. This tool will support the management of data in line with the DMP's goals and will aid in streamlining the workflow associated with data handling.

To promote the effective operation of the DMP and ensure its integration into the project's overall workflow, management actions include raising awareness among project partners. This will be achieved through targeted presentations and workshops focused on data management best practices.

1. Introduction

This document presents the initial Data Management Plan (DMP) for the WeForming project. This plan serves as the foundation for managing data within the project, with an emphasis on establishing core guiding principles for effective data handling. The DMP is intended to assist project members, partners, and practitioners involved in various stages of WeForming in systematically documenting, managing, and utilizing data. It emphasizes ethical usage and adherence to FAIR [2] principles throughout the project's lifespan and beyond.

1.1. Context and Scope

This DMP focus is set on the groundwork for the subsequent data management activities during the course of the project. Key references at include the WeForming project proposal document and other early-stage deliverables that outline use cases, requirements, and the system architecture. These documents provide a comprehensive overview of the project's scope and foundational information essential for the development of this plan. They offer insights into the initial requirements, anticipated data types, and ideas to inform and structure the work ahead.

1.2. Content and Structure

This section briefly outlines the content and structure of the DMP. The document commences with a discussion of the general guiding principles in Section 2, laying the foundation for the DMP's approach to data management; Section 3 introduces the WeForming solutions and instantiates the DMP framework accordingly. Section 4 delves into the data life-cycle and catalogue entry (i.e., the term refers to all the software and hardware solution that will be designed and implemented on the WeForming project) life-cycle, detailing how data will be managed from inception to disposal or archiving. Section 5 introduces the principles of FAIR data management, which are pivotal to the project's approach to handling data. In Section 6, the DMP expands its focus to include extended views on data management, encompassing considerations such as ethical aspects, data security, and other relevant issues. Section 7 presents the initial data management template and discusses the allocation of resources within the project for data management. This is followed by Section 8, which provides a preliminary introduction to the adopted issue tracking and proposed data management workflows. The document concludes with Section 9, which outlines the conclusions drawn so far and anticipates future work in the realm of data management for WeForming.

1.3. Target Audience

The target audience for this deliverable in the WeForming project encompasses a diverse range of stakeholders, including:

- **Partners and Advisory Group** within the WeForming project,
- **The European Commission (EC),**
- **Members of the European Union (EU Parliament,**

- **Other Horizon Europe projects,** particularly those related to energy and smart building initiatives (for clustering activities),
- **Organizations and experts** engaged in the WeForming case studies,
- **Other pertinent entities, both public and private,** which may include associations representing stakeholders relevant to the project's scope and objectives.

2. Guiding Principles

In accordance with the Horizon Europe framework [3], a Data Management Plan (DMP) is an integral part of effective data management for any research project. For the WeForming project, the DMP is designed to encompass the entire data management life cycle for all data that will be collected, processed, and/or generated. The aim is to ensure that our research data adheres to the principles of being FAIR. To achieve this, the DMP of WeForming will include detailed information on the following aspects:

- **Data Collection, Processing, and Generation:** Specify what types of data will be collected or generated, and how this data will be processed. This will encompass data from various sources, including but not limited to, energy consumption data from buildings, operational data from iGFBs, and IoT sensor data.
- **Methodology and Standards:** Outline the methodologies and standards that will be applied to ensure consistency, quality, and compliance in data handling. This will include protocols and standard data profiles (IEC, ETSI) for data collection, storage, and analysis, as well as standards for data quality and interoperability.
- **Data Sharing and Open Access:** Define the project's approach to data sharing and open access. This includes determining which data will be made publicly available, under what conditions, and through what platforms or repositories, ensuring compliance with ethical and legal standards.
- **Data Handling During and After the Project:** Describe the processes for handling and maintaining research data both during and after the completion of the WeForming project. This includes procedures for data use, access control, and ongoing management.
- **Data Curation and Preservation:** Elaborate on how data will be curated and preserved over time. This will include strategies for long-term data storage, data archiving, and ensuring continued data access and usability beyond the lifespan of the project.

These guiding principles set the stage for a comprehensive data management approach in the WeForming project, focusing on ensuring that all data-related activities are conducted ethically, responsibly, and in line with the best practices established by the Horizon Europe framework. Detailed implementation of these principles will be elaborated in Section 4, discussing the project data and catalogue entry life-cycle, and further operational details will be provided in Section 8, where the DMP will be fully presented.

2.1. Data Protection Legislative Framework

The WeForming consortium is acutely conscious of the ethical implications inherent in its research activities and upholds the ethical rules and standards set by Horizon Europe, as well as those outlined in the Charter of Fundamental Rights of the European Union. Wherever necessary, WeForming adheres strictly to both national and international laws, including:

Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016, regarding 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),**
- **The Directive on Protection of Privacy in the Telecommunication Sector (97/66/EC),**
- **The Universal Declaration of Human Rights,**

- **The Convention 108 for the Protection of Individuals** concerning Automatic Processing of Personal Data.

Additionally, WeForming aligns with

- **Article 19, 'Ethical Principles,' of Regulation No. 1291/2013/EC of the European Parliament and of the Council,**

which articulates the foundational principles of Ethics in research under Horizon Europe. This comprehensive approach ensures that all WeForming activities are conducted with the highest ethical integrity and respect for fundamental human rights.

3. Initial and Generic Data Items

In alignment with the WeForming project's objectives, the foundation for our data management strategy is outlined in the project's foundational documents, notably the GA. This document provides critical insights into the system architecture and the intended use cases for the project. It serves as the starting point for identifying and cataloguing the various types of data that the WeForming project will handle.

The preliminary data outlined in these documents represents the nucleus of our data management datasets. These data items are categorized as generic types, expected to be generated or utilized across the various activities of the WeForming project. Each category encapsulates a range of data types relevant to the development and demonstration of innovative solutions for energy management in buildings, including but not limited to:

- **Operational Data from iGFBs:** This includes data on energy usage, efficiency metrics, and operational performance of iGFBs within the multi-energy ecosystem.
- **Digital Operation Management Data:** Data related to the digital management of building operations, maintenance activities, and energy optimization strategies.
- **IoT Sensor Data:** Information collected from IoT devices deployed within the iGFBs, focusing on environmental parameters, energy consumption, and other relevant metrics.
- **Energy Ecosystem Interaction Data:** Data that encapsulates the interaction of buildings with the broader energy ecosystem, including energy market data, multi-user engagement metrics, and multi-sector energy usage patterns.

These data types, their sources, and methodologies for collection and processing constitute the key elements that the Data Management Plan (DMP) of WeForming will address. The comprehensive management of these data sets throughout their life-cycle – from collection and processing to sharing, archiving, and eventual disposal – will be elaborated in subsequent sections of the DMP.

3.1. Data Management Implementation

Table 1 is a conceptual alignment (with the project's focus on energy efficiency, smart grid technologies, and the involvement of various stakeholders in the energy building value chain) and may require further details or adjustments based on the specific operational nuances as the project progresses. The table takes into account the need for both public sharing of results and the protection of sensitive data, reflecting the project's emphasis on security, interoperability, and standardization (T4.2 -T4.5).

As for privacy and data protection, considering the project's focus on buildings and energy ecosystems, there is a heightened need for ensuring data privacy, especially with data pertaining to individual buildings or clusters. Therefore, anonymization and pseudonymization practices are essential for datasets containing sensitive information, aligning with the project's commitment to data security and privacy.

For scientific publications, open access is encouraged to maximize the impact and reach, in line with the communication and dissemination strategies.

TABLE 1: DATA MANAGEMENT IMPLEMENTATION

Dataset	Classification	Archiving	Safety and Security	FAIR	Privacy & Data Protection
WeForming Public Deliverables	Public	Each public deliverable will be openly accessible on the WeForming website and archived internally.	N/A	Deliverables available on the project website with meta-data	N/A
WeForming Open-Source Software Components	Public	Source code shared on GitHub or similar repositories.	N/A	Source code in GitHub or similar public repositories	N/A
WeForming Scientific Publications	Public	WeForming website and public repositories like Zenodo.	N/A	Publications indexed with appropriate meta-data	N/A
WeForming Pilot Data	Mixed (Public/Confidential)	Near real-time data stored on project DB; Asynchronous and sensitive data stored on premises.	Secured under cloud provider or individual partners' security policies.	Data translated to standard models, compliant with FAIR principles for public data. Stored in designated Data Brokers.	Anonymization or pseudonymization where applicable.
WeForming Internal Documents	Confidential	Stored in secure internal repositories, access restricted to project members.	Access controlled and encrypted storage.	N/A	Strict data protection and confidentiality protocols.

The WeForming Reference Architecture (RA) will be designed to facilitate the creation of an advanced digital platform for the efficient and intelligent management of energy within buildings. This architecture is the backbone of the WeForming Digital Platform, fostering effective and secure data sharing among stakeholders, addressing all needs to support upcoming data marketplaces in the smart building sector:

(A) Data Interoperability: An open-source, standardized, and domain-agnostic architecture, based on the Smart Grid Architecture Model (SGAM), ensures seamless interoperability of data across various systems within the smart building and energy management ecosystem.

(B) Data Sovereignty and Trust: Leveraging the International Data Spaces Association (IDSA) and FIWARE Conceptual Architectures, WeForming ensures not only open-source implementation for technological interoperability but also strengthens trust, sovereignty, and governance in data management. This

approach is particularly adapted to the unique requirements of IGFBs and their interaction with existing energy infrastructure.

(C) Data Value Creation: Embracing standardization initiatives, WeForming utilizes Open APIs, in line with industry standards such as NGSI-LD and/or the Context Information Management (CIM) API by ETSI ISG CIM, enhancing Context Information Management capabilities in the realm of intelligent buildings.

The utilization of W-IBRA will enable:

- **(i)** definition and categorisation new types of data assets specific to energy-efficient buildings;
- **(ii)** registration of offerings, which includes detailing asset descriptions, data models, access points, and the terms and conditions for data exchange, encompassing Service Level Agreements (SLAs), legal stipulations, and pricing models;
- **(iii)** facilitation the exploration, navigation, and discovery of existing offerings based on tailored criteria.

These functionalities will ensure that all data management activities within the project adhere to the FAIR principles tailored to the specific needs of energy-efficient building management.

4. Data and Catalogue Entry Life-cycle

Understanding the life-cycle of data and (open services) catalogue entries is crucial for the effective implementation of the Data Management Plan (DMP) in the WeForming project. This life-cycle elucidates the journey of data from its origin through various phases of handling, ultimately defining how it is managed and utilized within the project. The Catalogue term refers to all the software and hardware solution that will be designed and implemented on the WeForming project

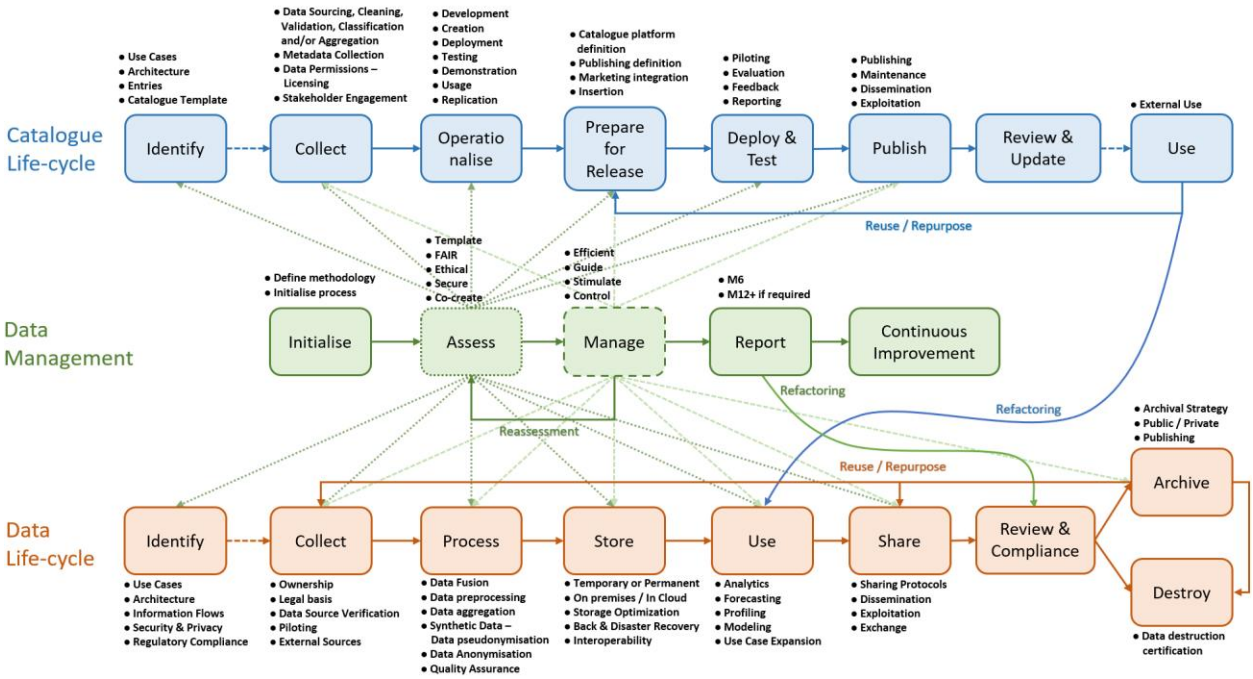

FIGURE 1: PROJECT DATA AND CATALOGUE LIFE-CYCLE AND THE DATA MANAGEMENT PROCESS.

Choice of Colors for each one of the horizontal processes of Figure 1: The blue color represents reliability and trustworthiness, qualities associated with a catalogue that serves as a reference. The green color symbolises growth, harmony, and safety, that aligns with managing data in an ethical and secure manner. Finally, the orange color is energetic and signifies creativity and enthusiasm, fitting for a process that involves transformation and the dynamic use of data.
The dashed arrows suggest a more flexible progression (for example from identification to collection, as in both life-cycles, "Identify" involves understanding what needs to be collected and actual collection might depend on factors like resource availability, compliance checks, or stakeholder inputs).
The "refactoring" arrows signify a feedback loop where the learning from one step informs improvements in another, encouraging continuous evolution of data practices to maintain relevance and effectiveness. Similarly, for the "Reuse/Repurpose" and rest of the arrows.

Figure 1 represents three separate but related process flows: the Catalogue Lifecycle, Data Management, and Data Lifecycle. The section that follows provides a brief explanation to familiarize the reader with the illustration and facilitate the general understanding of the utilised data management processes. The

content that follows in this section outlines the different stages of the processes and then includes "GA reflection" sections that specifically connect these stages to relevant activities and deliverables in the WeForming project as described in the GA. These reflections help to contextualize the general process steps within the specific scope and objectives of the project, showcasing how each stage of data management and processing is being actualized in the particular context. This approach aligns theoretical process steps with practical implementation but demonstrates how each phase is relevant to and integrated within the broader framework of the project's objectives and tasks.

4.1. Catalogue Lifecycle

The catalogue entry life-cycle in WeForming is designed to be dynamic, evolving with the project's progress and technical advancements. It serves as a vital component in the project's data management strategy, ensuring a systematic and efficient approach to handling the wealth of information generated and used in the project. The catalogue entries play a vital role as they encompass comprehensive information about the solutions and data sets being developed. The life-cycle of catalogue entries is closely aligned with the data life-cycle and is instrumental in planning, executing, and disseminating information both within and outside the project scope. The steps include:

Identify: This step involves defining the scope and structure of the catalogue, including use cases, architecture, catalogue entries, and templates. This is used as a comprehensive and logical starting point.

- **GA reflection:** This stage involves identifying what constitutes a relevant catalogue entry. The identification process aligns with deliverables and use cases as specified in various Work Packages, particularly in WP2 which deals with stakeholders' requirements and system specifications, and it feeds into this plan's assessment phase (referenced in T1.4).

Collect: Gathers necessary information and utilities for the catalogue. In this step a multifaceted approach is taken to gather the necessary data that forms the backbone of the catalogue. This begins with Data Sourcing, where diverse sources are tapped into to gather the required information. Following sourcing, Data Cleaning and Validation are conducted to ensure the information is accurate, free of duplicates, and relevant. Data Cleaning involves removing inconsistencies and errors, while Validation ensures that the data meets the pre-defined quality standards and criteria. Subsequently, Data Classification organizes the collected data into various categories, making it easier to manage and retrieve. This step aids users in navigating the catalogue more efficiently, as data is systematically grouped by type, source, relevance, or other specific criteria. Alongside, Data Aggregation might be necessary, where data from various sources is combined to create a more comprehensive set of information. Meta-data Collection is another critical step, wherein descriptive data about the collected information is amassed. Meta-data serves as a guide, detailing the nature, structure, and scope of the data, which is paramount for users to understand its context and potential applications. A significant legal aspect involves Data Permissions and Licensing, where the legal right to use the data is secured, and any associated terms are complied with. This step is crucial to ensure the catalogue's data is used ethically and within legal bounds. Lastly, Stakeholder Engagement is an interactive phase where feedback from potential users and stakeholders is sought. Their insights help in refining the data collection process and ensuring the resulting catalogue meets the practical needs of its users. This engagement can also lead to the identification of additional data sources or highlight the need for specific data types within the catalogue. Overall, the "Collect" step requires careful attention to detail

and a methodical approach to ensure that the catalogue is not only comprehensive and legally sound but also aligned with user requirements and expectations.

- **GA reflection:** The collection of entries is a multi-WP effort. For example, WP3 focuses on building blocks and technological enablers, contributing significantly to the collection of entries related to energy optimization solutions and the transformation of passive buildings to iGFBs.

Operationalise: This involves the practical development and deployment of the catalogue, including testing and demonstrating its use. In this step, a seamless transition from conceptual design to a working catalogue is achieved. The process begins with the Development phase, where the technical backbone of the catalogue is created. Next, the Creation phase involves populating the catalogue with carefully curated content. Deployment then takes the catalogue live, ensuring it is fully functional for users. Once deployed, the catalogue is rigorously Tested for any issues that might impede user experience. It is then Demonstrated to stakeholders for feedback and user training. In the Usage phase, real-world use informs further refinement of the catalogue. The final Replication phase identifies and codifies successful strategies for future initiatives. Throughout this process, from development to user engagement, the catalogue is transformed into a robust and valuable resource for its intended audience.

- **GA reflection:** This phase is concerned with the digital interoperability framework and integrated solutions, including the design and development of open service catalogues and data space connectors, aligning with the objectives of WP4. These operational aspects make catalogue entries functional, supporting user engagement and community services.

Prepare for release: This step is about defining the catalogue platform and preparing for publishing. It encapsulates the final preparations for launching the catalogue. It's where the chosen hosting platform is tailored to ensure a seamless user experience, the release mechanisms are defined to allow straightforward access to content, and marketing strategies are woven in to maximize the catalogue's reach. The culmination of this step is the actual placement of the catalogue on the selected platforms, readying it for discovery and use by the target audience.

- **GA reflection:** This step involves building the catalogue and integrating related services, potentially interacting with marketplaces or platforms. This phase also corresponds to tasks in WP4, which include the development of service catalogues and interoperability middleware, essential for the digital platform's seamless functioning.

Deploy & Test: This step includes piloting, evaluation, and feedback, which should ideally come before publishing. It begins with Piloting, putting the catalogue into operation in a controlled environment to test its functionality. During Evaluation, the catalogue's performance is assessed against objectives. Feedback is gathered from pilot users to identify any issues or areas for improvement. Finally, Reporting involves documenting the results and feedback from the pilot to guide any necessary adjustments before the full rollout. This step ensures the catalogue is robust and user-ready upon launch.

- **GA reflection:** Deploying and testing on catalogue entries fall within the domain of WP5, which is dedicated to demonstrating technologies and scenarios across various environments. Feedback from these phases is vital for the continuous enhancement of the catalogue entries.

Publish: This is about making the catalogue available and maintaining it. This step is where the catalogue is released to the public. Once live, Maintenance is critical to ensure ongoing functionality and relevance. The Dissemination process involves actively sharing the catalogue to reach a wider audience. Lastly, Exploitation focuses on leveraging the catalogue's assets to their fullest potential,

whether for commercial purposes, educational use, or further research. This step is key for the catalogue to be utilized effectively and sustain its value over time.

- **GA reflection:** Making catalogue entries accessible aligns with WP6, which focuses on evaluation, assessment, replication, and scalability potential. The entries might be published on platforms like GitHub or the project's website, offering accessibility to a broader audience.

Review & Update: Allows for iterative improvements based on user feedback.

- **GA reflection:** Tasks across WP2 and WP4 contribute to this phase. Feedback from stakeholders and end-users, derived from use-case analyses and system specifications, informs continuous updates and improvements in the catalogue. The review of deliverables D2.1, D4.2 enables the identification of areas where the catalogue needs updates, ensuring it accurately reflects the latest project developments and stakeholder requirements. Ensuring compliance with regulatory changes, as detailed in WP1 is a key aspect of this phase. As regulatory landscapes evolve, particularly in the energy sector, the catalogue must be reviewed and updated to maintain compliance. Finally, continuous technological advancements in energy management and smart building solutions, explored in WP3 necessitate regular updates to the catalogue to incorporate new technologies, tools, and methods.

Use: The final step, where the catalogue is put to external use, by end-users or other stakeholders.

- **GA reflection:** The final utilization of the catalogue entries, potentially by external partners, ties in with the overall objectives of the project, which include creating sustainable and efficient building energy management systems. The feedback obtained during this phase will inform future refinements and developments. This phase finds its practical application primarily in WP5 where the catalogue's contents are actively used in real-world demonstrations (D5.1, D5.2). Also, the catalogue's resources, including best practices, technologies, and methodologies developed in WP3, are actively employed in this phase. Finally, in line with the objectives of WP6, the data and insights gathered from the catalogue are used for comprehensive impact assessments, supporting informed decision-making regarding scalability and replication potential.

4.2. Data Management

Initialize: This step is defining the methodology and starting the process. This step sets the foundation for how data will be handled. It involves defining a clear Methodology that outlines the protocols, standards, and tools that will be used throughout the data management lifecycle. Once the methodology is established, the process is officially initiated, setting the data management activities into motion. This initial step is crucial as it dictates the framework within which the subsequent data management tasks will be executed.

- **GA reflection:** Aligns with WP1, specifically Task 1.4 (Data Management Planning). This initial stage sets the data management foundation, as outlined in D1.1.

Assess: This step involves reviewing the established Template to ensure that it effectively captures all necessary data attributes and supports the FAIR principles. Attention is given to Ethical practices, confirming that data management upholds the moral obligations towards data subjects' privacy and consent. The security protocols in place are rigorously examined to ensure that data remains Secure against breaches and leaks. Furthermore, this step includes an appraisal of the co-creative processes used in developing and managing data. This ensures that stakeholder collaboration is not only encouraged but also results in meaningful contributions to the project's data management strategies. This phase is crucial for maintaining a data management system that is robust, ethical, aligned with best practices, and dynamically improved through stakeholder engagement.

- **GA reflection:** This step reflects the ongoing assessment in WP6, particularly Task 6.1 (Impact Assessment Methodology). Regular assessments ensure the data management strategies align with Deliverables D6.1 to D6.3, which focus on various impact assessments.

Manage: This step includes guiding and controlling data management practices. This is where the active governance of data takes place. This phase is all about ensuring efficiency in handling data, from storage to retrieval. It involves guiding all stakeholders on best practices for data usage and maintenance. The process also aims to stimulate productive use of the data by encouraging innovation and effective application. Lastly, control mechanisms are implemented to oversee data access and modifications, ensuring compliance with policies and maintaining the integrity of the data throughout its lifecycle. This step is essential for sustaining a high-quality, reliable data management system.

- **GA reflection:** Corresponds with WP4 especially Task 4.5 (Security, Privacy, and Data Sovereignty), overseeing the operational management of data. This step ensures compliance with WP4's objectives, maintaining data integrity and security.

Report: Reporting at specified milestones (M6, M12 if required, etc.) ensures that data management practices are reviewed and updated regularly. This step could also encompass continuous monitoring and reporting rather than fixed milestones.

- **GA reflection:** Mirrors the structured reporting in WP1, particularly through Deliverables D1.2 (Periodic Technical Reports) and D1.3 (Final Report), providing documentation and updates on data management practices at key project milestones.

Continuous Improvement: Continuous Improvement is based on reporting outcomes and evolving project needs. This step solidifies the cyclical nature of the Data Management process, ensuring that the practices are not static but evolve over time. This step is about taking the insights gained from reporting at specific milestones or continuous monitoring and using them to make data management practices more effective. This includes refining methodologies, updating security protocols, enhancing data quality, and ensuring that the data management strategies remain aligned with the latest technological advancements and compliance requirements. It's a step that encourages a proactive stance towards growth and adaptation. As data management is a dynamic field influenced by changing regulations, technological innovations, and evolving user needs, this ensures that the data management system is not only responsive to current requirements but also anticipates future trends and challenges.

- **GA reflection:** Addresses the iterative improvement approach in WP3, particularly seen in Task 3.5 (Design and Development of AI/ML Tools). Insights from earlier stages are used to refine data management practices, ensuring alignment with the evolving needs of the project as evidenced in Deliverables D3.1 to D3.3.

4.3. Data Life-cycle

The Data Lifecycle is essentially about managing data from its inception through its active use and eventual archiving or destruction. It's a continuous, iterative process, which ensures that data is handled efficiently and ethically throughout its existence. The Data Lifecycle part of the process is critical because it covers the entire journey of data from its inception to its eventual archival or deletion. This lifecycle ensures that data is treated as a valuable and sensitive asset throughout its lifespan. Each step is designed to ensure that data is handled ethically, legally, and efficiently, maximizing its utility while minimizing risks.

Identify: Similar to the Catalogue Lifecycle, this involves defining the scope of the data collected, including legal and privacy considerations. It's well-structured and crucial for compliance. This step is pivotal as it sets the parameters for how data will be managed throughout its journey. This phase begins with defining Use Cases to understand the specific scenarios in which the data will be utilized. This informs the Architecture, structuring the data framework to support these use cases efficiently. Next, Information Flows are mapped out to determine how data moves through the system, ensuring a smooth transfer of information between processes. Security & Privacy are foundational concerns, addressed from the start by integrating safeguards that protect data against unauthorized access and breaches, while ensuring that personal information is handled respectfully. Finally, Regulatory Compliance is assessed to guarantee that all data handling practices are in line with current laws and regulations, laying the groundwork for a data life-cycle that is responsible, secure, and legally sound. This step is critical for establishing the integrity and usability of the data, dictating its value and reliability for future use.

- **GA reflection:** The data life-cycle in WeForming starts with the crucial phase of data identification. This is informed by the initial project documents, specifically the GA, which outlines the project's use cases and system architecture. These foundational elements set the stage for identifying the types of data the project will engage with, including but not limited to:
  - **Use Cases:** The use cases of WeForming, essential for data identification and collection, are derived from various tasks and deliverables, notably from Deliverable D2.1 (SOTA analysis, barriers, regulatory framework, and end-users' requirements) and Deliverable D2.2 (WeForming Services Co-Creation, Business Models, Functional Specifications, and Reference Architecture). These documents provide a comprehensive analysis of end-users' requirements, state-of-the-art technologies in energy storage solutions, and the development of business models and functional specifications. The use cases define interactions among various stakeholders, particularly focusing on the integration of buildings with the energy grid and optimizing energy use in iGFBs.
  - **Architecture:** The WeForming architecture, progressively detailed in Deliverables D2.2, D2.3, and D2.4 (all detailing the Services Co-Creation, Business Models, Functional Specifications, and Reference Architecture), outlines the interaction of various components in the WeForming platform and is instrumental in the data life-cycle management, ensuring that data flows smoothly and securely across the system. These deliverables set the framework for encouraging interoperability services and standardized data integration between iGFBs and their energy ecosystems.
  - **Security & Privacy:** Security and privacy considerations are addressed in Task 4.5 (Security, Privacy, and Data Sovereignty) and are designed in line with tasks of WP1 (T1.3, T1.4). These elements ensure that all data collected and processed within the project framework adheres to strict security protocols and privacy norms, particularly in compliance with GDPR. The focus is on ensuring that data privacy impact assessments and security frameworks align with the cutting-edge technological enablers being developed, as detailed in D3.1, D3.2, and D3.3.
  - **Information Flows:** The documentation of information flows is crucial for data management and is elaborated in Deliverables D4.1, D4.2, and D4.3. These deliverables describe the integrated platform, including the catalogue of services, data space connectors, middleware, and the integrated solution, thereby detailing the flow of data through various components of the framework.

Collect: This step is about laying the ethical and legal groundwork for data collection, ensuring that the data's integrity is beyond reproach from the point of collection through its entire life-cycle. This involves the actual gathering of data from various sources, ensuring legality and ownership clarity.

Ownership must be clearly defined to establish who has the rights and responsibilities over the data. This is closely tied to the Legal basis, which ensures that data collection complies with applicable laws, such as GDPR for personal data. Data Source Verification is critical to confirm the reliability and accuracy of the data sources, which includes validating the credibility of the data and the reputation of the source. Piloting is a practical phase where data collection methods are tested on a smaller scale before full implementation, helping to identify potential issues early on. Lastly, External Sources may be approached to enrich the dataset, which could include public data repositories, commercial data providers, or cross-industry collaborations.

Following identification, the data life-cycle in WeForming encompasses:

- **GA reflection:** The collection of data is guided by specific use cases and system architecture as described in the GA and relates to the gathering of data essential for the project's objectives, involing several work packages and tasks; WP2 (WeForming stakeholders' requirements and system specifications) which is primarily focused on requirement engineering and specifications for software components developed in the project, WP3 (WeForming building blocks and technological enablers specification, design, and development facilitating the transformation of passive buildings to iGFBs) which includes tasks like extending proprietary digital platforms and tools to fit the WeForming Frameworks (Task 3.1), a task that is crucial for the collection and integration of data, and WP4 (WeForming Digital Interoperability Framework and Integrated Solution) which involves designing and developing open service catalogues (Task 4.1), which would also include the collection and cataloguing of services that active buildings can offer, contributing to data collection processes. The process is also informed by the planning and execution documentation of demos, as outlined in Deliverables D5.1 through D5.4. These deliverables document the preparatory activities, baseline scenarios, and performance measurement plans for the pilots, providing a framework for data collection during the demo operations. Finally, legal and ethical considerations, especially regarding personal data, are managed in line with WP1 (Project Management and Administration) tasks, ensuring compliance with GDPR and other regulations.

As to the complete life-cycle of data (from processing to destruction) it is overseen with a focus on multi-dimensional impact assessment as part of T4.5. The deliverables D6.1 through D6.3 play a key role in evaluating the data collected from various dimensions (technological, economic, societal, environmental), thereby influencing how data is used, shared, archived, and eventually, disposed. The following lines address the data life cycle process adopting a brief mapping to specific project elements, where applicable, in an exemplary fashion.

Process: Data processing steps including anonymization and pseudonymization are critical for privacy. This step is where collected data undergoes transformation to become more meaningful and secure. Data Fusion integrates different datasets to provide a comprehensive view. This is followed by Data Preprocessing, which refines the data, sorting and cleaning it for analysis. Data Aggregation then summarizes the information for easier interpretation. In terms of privacy, Synthetic Data and Data Pseudonymisation alter data to protect individual identities, while Data Anonymisation removes identifiable information entirely. Throughout this phase, Quality Assurance is critical, ensuring the processed data maintains high standards for subsequent use. This step is vital in converting raw data into reliable, insightful, and ethical assets for the project.

- **GA reflection:** The data processing phase involves steps such as data cleansing, pseudonymization, or anonymization. This phase is integral to WP4 (WeForming Digital Interoperability Framework and Integrated Solution), such as T4.4 and T4.5, where data is

prepared for analytics and decision-making tools with the objective to process and enhance utility while ensuring privacy and security.

Store: This step is about strategically securing the processed data. It involves deciding between Temporary or Permanent storage based on the data's relevance and lifecycle. The choice between On-premises or Cloud storage is made, considering factors such as accessibility, cost, and security. Storage Optimization ensures data is stored efficiently, maximizing space and resources. Backup & Disaster Recovery plans are established, safeguarding data against potential loss or corruption. Lastly, Interoperability is addressed to ensure that the stored data can be used across various systems and applications, facilitating collaboration and data exchange. This step is crucial for maintaining the integrity, availability, and usefulness of the data.

- **GA reflection:** After processing, data is stored in accordance with the requirements laid out in Task 2.7 (WeForming Operational Framework design – Building Interoperable Reference Architecture for optimized building-to-grid integration), which ensures data integrity and availability within the WeForming Middleware architecture (aligning with requirements from WP3).

Use: This step encapsulates the active application of stored data to derive value and insights. This involves deciding on the Temporary or Permanent use of the data depending on project needs. Analytics are applied to interpret the data, uncovering patterns and insights. Forecasting uses the data to predict future trends, aiding in strategic decision-making. Profiling helps to understand characteristics of different entities, while Modeling uses the data to create simulations or representations of reality. Use Case Expansion explores new ways the data can support additional scenarios or applications, broadening the impact of the data. This phase is essential for leveraging data to support informed decisions and innovations.

- **GA reflection:** During the usage phase, data becomes actionable insights through analytics, crucial for WP3's objectives (WeForming building blocks and technological enablers specification, design, and development), especially as seen in Task 3.5, which focuses on the Design and Development of AI/ML tools for the operation of iGFBs.

Share: In this phase, data is distributed beyond its initial creation and use environments. It begins with establishing Sharing Protocols, which set the rules and methods for how data is to be shared, ensuring it is done securely and ethically. Dissemination involves the broad release of data or findings to reach a wider audience, often through publications, databases, or digital platforms. Exploitation then takes the shared data and applies it practically, looking to extract value, whether for commercial gains, educational purposes, or policy development. Exchange denotes the reciprocal sharing of data between entities, which can facilitate new partnerships and collaborative opportunities. This stage is critical for multiplying the data's value through wider distribution and use.

- **GA reflection:** Controlled and secure data sharing protocols, especially for user-related data, are outlined in WP4, particularly Task 4.1, which deals with the Design and Development of an Open Services Catalogue and an App Store Towards Building Occupants, Managers, Grid Operators, and Markets. Findings are shared with stakeholders and the community, in line with dissemination activities in WP7 (Communication, Dissemination, Exploitation and Market Exploration). Sharing protocols ensure only anonymized or non-sensitive data is public, protecting confidentiality.

Review and Compliance: This step serves as a checkpoint before the data is archived or destroyed. In this phase, a thorough Review of the data is conducted to ensure that all handling, sharing, and usage have adhered to established guidelines and policies. This review process checks for compliance with various standards, including legal regulations, ethical norms, and organizational policies that govern

data use and retention. Compliance is particularly crucial, as it verifies that data management practices have been consistent with the latest laws and regulations. This stage helps identify any discrepancies or issues that need to be addressed before data can be confidently archived for long-term storage or securely destroyed. It's a safeguard that ensures responsible data stewardship and upholds the trust of data subjects and stakeholders. This proactive step demonstrates a commitment to due diligence and integrity in the data management process.

- **GA reflection:** This phase aligns with the compliance activities outlined in WP1 and the continuous monitoring and evaluation efforts in WP6. In WP6, data practices are regularly reviewed to ensure they align with the broader objectives and ethical considerations of the project, as reported in D6.1 through D6.3 (Impact Assessment Deliverables). As regards adherence this phase ensures that data handling is in line with objectives and use cases outlined in WP2 and WP3.

Archive: Data that has long-term value is archived. It's important to determine what data is worth preserving, balancing the costs of storage against the potential future value. This step focuses on the long-term preservation of data. This begins with developing an Archival Strategy, which determines how and where data will be stored for future access. Decisions are made regarding whether the archives will be Public or Private, balancing the need for openness with privacy and security considerations. Publishing the data involves making it available through institutional repositories, databases, or other platforms, ensuring that valuable information remains accessible for future research and reference. This phase is crucial in safeguarding the data's legacy, ensuring that it remains a resource for ongoing learning and discovery.

- **GA reflection:** Archiving strategies are developed in accordance with the broader data management objectives and principles of the project, as covered in Task 1.4 (Ethics exchange requirements specifications and data management) and WP6, which includes evaluation and assessment activities.

Destroy: Finally, data that is no longer needed or has reached the end of its retention period must be destroyed securely. This step is crucial for maintaining privacy and compliance with data protection regulations. This step is often guided by a Data Destruction Certification, which is a formal process that verifies the data has been irreversibly destroyed in compliance with legal and regulatory standards. This certification provides documented assurance that the data was disposed of responsibly, preventing any potential data breaches or unauthorized recovery after its intended life-cycle has concluded. The inclusion of this certification ensures that the destruction of data is performed transparently and can be audited if necessary.

- **GA reflection:** The destruction, or secure disposal, of data no longer needed or that has fulfilled its purpose, including personal data, is performed in compliance with GDPR and ethical guidelines. Task 1.4 manages these aspects to ensure that all data is handled properly throughout its lifecycle.

5. FAIR Data Management

In the WeForming project, adherence to FAIR data management principles is central to ensuring that the data collected, processed, and generated are findable, accessible, interoperable, and reusable. These principles are guided by the Horizon Europe framework and are reflected in our data management strategies.

- **Findable:** In WeForming, data findability is ensured by employing well-defined meta-data and standard identification mechanisms. Metadata is designed not only to meet the immediate needs of the project but also to facilitate potential future re-use of the data. Standardized meta-data formats and data formats are used wherever possible, enhancing the ease of data discovery. The data management system incorporates a clear versioning system to track multiple revisions of data and meta-data, as highlighted in the project's Data Management Plan (D1.2).
- **Accessible:** The WeForming project commits to making data openly available as a default, subject to privacy and confidentiality considerations. Data intended for open access is described in terms of its publishing method, favoring open, standard, and certified repositories. In cases where data cannot be openly shared, reasons for such restrictions are clearly documented. Additionally, any specialized software required for data access is preferably open-sourced, and when data is published, it is done so under clearly defined open licenses.
- **Interoperable:** Our approach to interoperability involves using open and standardized data formats that are compatible with commonly used software applications. This ensures the seamless exchange of data with other researchers and organizations, fostering a collaborative and productive research environment. The interoperability of data is a key consideration in the development of the WeForming integrated platform (D4.1, D4.2, D4.3), ensuring that data can be easily integrated and utilized across different systems and components of the project.
- **Reusable:** The reusability of data in WeForming is maximized through careful selection of data licenses, consideration of the timeframe and actuality of data release, and a focus on data quality and usability. The goal is to ensure that data remains relevant and valuable for future research or application, beyond the scope of the current project. The reusability aspect is considered in the context of the project's various deliverables and work packages, ensuring that data generated can serve as a valuable resource for future initiatives and research.

5.1. Making Data Findable

In the WeForming project, all data management activities, such as storage, processing, and mutual sharing of information among project participants, will be facilitated through a dedicated platform, distinct from ProofHub, while ensuring public engagement with the project's findings via the official project website. All data will be archived in the project coordinator's private cloud storage system for at least five (5) years post-project completion, with a possible extension of up to two (2) years upon request. Confidentiality, record-keeping, and impact evaluation will follow the same duration.

A **naming convention** will be adopted that succinctly details the content, data-collecting institution, and publication month. **Version control** will be applied where necessary, especially in instances of data withdrawal, incorporating version numbers into file names.

Technical steps will be implemented to ensure data anonymity, with identifiable information leading to real names not being distributed. The project will handle data primarily from building energy management systems, collected or generated from various sources like monitoring sensors, smart devices, and user interactions throughout the building management lifecycle.

The data sovereignty module will augment existing frameworks for data handling and distribution, establishing a data space with open data models, standardized APIs, and effective data sharing policies, all aligned with FAIR principles.

5.1.1. DATA DISCOVERABILITY

Data generated within the scope of WeForming will be traceable and locatable through unique identification protocols. Files will be distinctly identified using standardized naming conventions and file versioning procedures. Meta-data annotations will adhere to standards like the Dublin Core generic meta-data standard [4], to broadly describe research data.

To align with FAIR principles, (meta)data will:

- Be associated with a **unique and permanent global identifier,**
- Include **comprehensive meta-data** for complete data interpretation,
- **Be catalogued in a source facilitating easy searching.**

5.1.2. DATA IDENTIFICATION MECHANISMS

All documents will carry the project's name, alongside a unique, enduring identifier for the document type and number, as provided by the coordinator for submission to the European Commission (EC). Each document's identification will include the task or deliverable number corresponding to it, followed by a succinct title of the activity or deliverable.

For academic article dissemination, Digital Object Identifiers (DOIs) will be issued by the publisher. Other forms of written work, such as reports and policy recommendations, will receive DOIs via the repository where they are archived (e.g., Zenodo).

5.1.3. NAMING CONVENTIONS

To enhance data searchability and discoverability, every data set (datasets, deliverables, etc.) in the project will follow a consistent naming protocol, incorporating a version control table. Naming guidelines will ensure identifiers are meaningful yet concise, avoiding language-specific characters or non-alphanumeric symbols, and appending a two-digit numerical suffix for new document versions.

For deliverables, the naming format will be: **Project's name - Dx.y - [Name of the deliverable as described in the DoA]**, where x is the work package number and y is the deliverable number within that package. For datasets: **Project's name - WP [Work Package number] P [Pilot number; pilot activity number] - T [Task number; description of the activity**].

The project will use user-friendly search keywords to facilitate effective data reuse by stakeholders. Keywords will reflect relevant subject matters, including energy efficiency, intelligent building

management, smart grids, interoperability, sustainable architecture, and more, accurately mirroring dataset content without redundancy.

Typically, the keywords will encompass terms related to the core topics of WeForming, including energy efficiency in buildings, intelligent energy management, smart building technology, IGFBs, sustainability in architecture, digital twin technology, smart grids, data interoperability, and cutting-edge business models for smart cities. Additional project-specific keywords, such as WeForming, Horizon Europe, and various relevant standards and strategic initiatives, will also be employed. These keywords will be carefully chosen to reflect the essence of the datasets, ensuring relevance and avoiding terms that are infrequently used or peripheral to the main content of the project.

5.1.4. APPROACH TOWARDS SEARCH KEYWORDS

For documents pertaining to the project activities, standardized templates agreed upon by the consortium will be utilized. These templates will include a designated section for keywords, specifically designed to enhance the findability of the documents. This approach ensures that all project-related materials are easily searchable and accessible, aligning with the project's commitment to efficient information management and dissemination.

5.2. Open Data Accessibility

Subject to ethical considerations and participant consent, data in the project will be as accessible as possible. WeForming will utilize a secure identity management system through its data sovereignty module for the certification, identification, authentication, and authorization of organizations and individuals (i.e., data-space aligned principles [5]). In addition, trusted data exchange mechanisms will ensure that data is not only accessible (**attainable**) but also **reliable**.

Regarding scientific data production in WeForming, the focus will include: **(i)** Advanced learning data models that offer deeper insights and facilitate the development of data-driven services for efficient building energy management. These services will exploit the rich data from iGFBs ensuring interoperability with energy network operators. **(ii)** The integration of cutting-edge technologies in smart building management systems. This involves implementing innovative control, operation, and management methods, leading to optimized energy efficiency and enhanced integration of Renewable Energy Sources (RES) into building ecosystems.

Dissemination levels for WeForming reports will vary. Public reports will be disseminated through public platforms, while confidential reports will remain within the consortium. Scientific publications will be made freely accessible and uploaded to public repositories. Data sharing among partners will occur through a designated shared space. Some research data, particularly those related to demo users, will be sensitive due to privacy and data protection considerations and will therefore be handled confidentially or anonymized before being made accessible. Data will be made available wherever feasible, subject to ethical guidelines and participant agreement. However, due to the nature of data collected in WeForming, especially personally identifiable information, there will be instances where it is challenging to release such data.

Table 2 will outline the data that are: **a)** produced and used in the project, and **b)** slated to be made openly available.

TABLE 2: DATASET AVAILABILITY

#	Data Type	Data openly available (y/n)	Justification
1	Stakeholder contacts collection	No	The data related to the stakeholder's contacts will not be published as primary data due to privacy and security concerns.
2	Public security services collection	Yes	The data related to public security services will be made openly available because it pertains to general public services information that is not sensitive or personal. This data is crucial for understanding and analysing the public safety aspects related to energy management in iGFBs and its availability will support broader research and development efforts in this field.
3	Digital security solutions collection	Yes	Information regarding digital security solutions will be publicly accessible, as it contributes to the broader knowledge base of digital security practices within smart building management systems. This openness fosters collaborative development and innovation in the field, allowing for the sharing of best practices and encouraging standardization across similar projects.
4	Quantitative survey data	Yes	The accessibility of this data is essential for the broader understanding and validation of the project's findings and methodologies. It aligns with WP6's focus on the evaluation, impact assessment, and the potential for replicability and scalability of the project's results. Open access to quantitative survey data enables peer review, replication studies, and contributes to the scientific community, facilitating broader discussions and advancements in smart building energy management.
5	Expert interview data	No	The data from the expert interviews (recordings, protocols and transcriptions) will not be published as primary data due to privacy and security concerns. Anonymization is not considered as an alternative, because the sample size allows drawing conclusions on the respondents.
6	Focus groups data	No	The data from the focus groups (recordings, protocols and transcriptions) will not be published as primary data due to privacy and security concerns. Anonymization is not considered as an alternative, because the sample size allows drawing conclusions on the respondents.
7	Workshops data	No	The data from the workshops (recordings, protocols and transcriptions) will not be published as primary data due to privacy and security concerns. Anonymization is not considered as an alternative, because the sample size allows drawing conclusions on the respondents. Webinars/workshops may be recorded and posted in on our YouTube channel. In this case, - at the beginning of the webinar there will be a disclaimer asking everyone to accept that it is being recorded. The videos will be edited afterwards, removing the external stakeholder comments

			and questions, and leaving only the training material available on YouTube channel. To achieve broader project exposure through workshops, relevant stakeholders will be considered to be invited. The privacy methodology related to the collection of stakeholder emails, through WP2 partners through surveys, asking for their scope and interests within the energy ecosystem; as a matter of fact, inviting them to the webinars and co-creation activities targeted according to their answers. The emails to be collected will be maintained only by the F6S and the partner in charge of the survey have access to these emails.
8	Validation cycles data	No	The data from evaluation survey will not be published due to privacy and security concerns. Anonymization is not considered as an alternative, because the sample size allows drawing conclusions on the respondents.

As it was indicated by the aforementioned information, the following data sets will be made openly accessible: Data type **#2** (Public security services collection), **#3** (Digital security solutions collection) and **#4** (Quantitative survey data). Table 3**Error! Reference source not found.** describes the accessibility details of these particular datasets.

TABLE 3: DATASET ACCESSIBILITY

#	Data Type	Level of accessibility	Type of availability and required software tools	Information on meta-data and additional data information
2	Public security services collection	Public	Filterable and searchable database; can be accessed with a state-of-the-art web browser	No meta-data needed; additional information will be provided on the platform
3	Digital security solutions collection	Validated professionals	Filterable and searchable database; can be accessed with a state-of-the-art web browser	No meta-data needed; additional information will be provided on the platform
4	Quantitative survey data	Public	Cleaned primary data; can be accessed with SPSS, PowerBI, Excel or any similar data analysis tool.	No meta-data needed; additional information will be provided on the platform

5.2.1. METHODS OR SOFTWARE NEEDED TO ACCESS THE DATA

In the WeForming project, no specialized software tools will be required for accessing the data. To ensure ease of use and long-term accessibility, anonymized datasets will be stored in universally accessible formats such as Word, PDF, or Excel. This approach is designed to facilitate data exploitation and accessibility across a wide range of users.

5.2.2. DATA DEPOSIT, ASSOCIATED META-DATA, DOCUMENTATION & CODE

For WeForming, each Work Package (WP) leader will bear the responsibility of depositing and securing the data generated from their respective project activities. In addition to storage within the WP leaders' systems, an additional copy of all final deliverable forms will be maintained on the coordinator's

designated project management platform (akin to ProofHub) and, for certain maintenance scenarios, in their cloud storage solution (comparable to Dropbox). This dual-storage approach ensures data integrity and availability throughout and beyond the project's lifespan.

5.3. Data Interoperability

The principle of interoperability necessitates that data be machine-readable and terminology consistent. In line with this, the WeForming project employs open-source, standardized, and domain-agnostic Open APIs to ensure seamless data interoperability across various systems in the smart building and energy management domain. The project will aim to built upon the foundation of the Service Catalogue and the App Store both acting as Vocabulary Provider to enhance the interoperability among WeForming stakeholders.

The project will establish appropriate protocols for the creation of data and meta-data, along with the development of relevant vocabularies. For WeForming, achieving data interoperability is essential, as the project focuses on creating semantic data models and ontologies tailored to intelligent energy management in buildings. The evaluation and development of standards and interoperable data models will be a key focus within the project, particularly in WP4, to facilitate the effortless sharing and integration of data from diverse building management and energy systems. WP4 aims to develop a fully interoperable, interconnected, and trusted system in alignment with the project's building reference architecture. This includes designing and implementing open service catalogues, app stores, open data space connectors, interoperability middleware, and system integration. The focus is on enabling semantic and service interoperability, secure data exchanges, and integral methodologies and digital solutions to support demonstrators' applications.

5.3.1. DATA ASSESSMENT INTEROPERABILITY

The sophisticated Data Governance Middleware in WeForming integrates the Function, Information, and Communication Layers as per the Smart Grid Architecture Model (SGAM), establishing a dynamic framework for its digital platform. Within this structure, partners are tasked with maintaining data in an accessible and comprehensive format, tailored to the evolving requirements of stakeholders who may utilize, integrate, or leverage the data produced during the project. Regular assessments of data interoperability will be conducted, ensuring that the project's data continuously aligns with the specific needs of various scenarios, including diverse data infrastructures and the unique objectives or interests driving data use.

5.3.2. INTERDISCIPLINARITY AND TRANSDISCIPLINARITY

WeForming is dedicated to promoting knowledge collaboration among experts from varied disciplines, breaking down the barriers of traditional academic fields. The project inherently adopts an interdisciplinary approach, uniting diverse scientific expertise to forge an efficient, integrated, and intelligent energy management system for iGFBs. Table 4 outlines the array of disciplines involved and the corresponding Work Package(s) where these are integrated into the data management of WeForming. The consortium of partners brings together a broad spectrum of data generation and processing capabilities, leveraging their knowledge and expertise to ensure thorough coverage essential for the successful execution of all project activities.

TABLE 4: DATASET KNOWLEDGE AREAS

Discipline(s)	WP(s)
Sociology (for ensuring consumer acceptance and engagement)	2, 7
STEM i.e., Physics, Mathematics, Data and Computer science (along with specific areas of science related to energy, financial, engineering, industry, water, agriculture, and mobility sectors for the design of the tools and systems)	2, 3
STEM for the Demos purposes	5
Economics (for the development of viable business models)	2
Economics (for the exploitation plans)	7
Education science (for organizing and implementing workshops / capacity-building activities)	7
Political sciences (for better understanding of regulatory measures)	2, 7
Decision-making process and pursuing collaborative multi-level governance	1, 7

For the reader's convenience, below stands a quick reference bullet-list providing the descriptive titles of WeForming's WPs, as follows:

- **WP1:** Project Management and Administration
- **WP2:** WeForming stakeholders' requirements and system specifications
- **WP3:** WeForming building blocks and technological enablers specification design and development
- **WP4:** WeForming Digital Interoperability Framework and Integrated Solution
- **WP5:** WeForming demonstration across technologies and scenarios
- **WP6:** Evaluation and assessment replication and scalability potential
- **WP7:** Dissemination exploitation standardisation and impact outreach

6. Extended Views

In WeForming, several extended views are integral to the data management process, encompassing data security, privacy, ethical considerations, and compliance with various policies and regulations.

Data Security and Privacy: Ensuring the security and privacy of data collected, stored, processed, and used in the project is paramount. The foundational security and privacy requirements are set out in deliverables D2.2 to D2.4, addressing the functional specifications for interoperability and standardized data integration. These deliverables lay the groundwork for the data protection strategies within the project. Additionally, the Data Management Plan, as outlined in Deliverables D4.1 to D4.3, covers the security aspects of data sharing. Tasks under Work Package 4, especially T4.5 (Security, Privacy, and Data Sovereignty), directly contribute to the development and enforcement of these security measures, ensuring that data handling within the project's ecosystem is compliant with the highest standards of data protection.

Ethical Aspects: The WeForming project places significant emphasis on ethical considerations, especially in terms of data privacy. In alignment with the project's ethical guidelines and deliverables (T1.4), no personal data is shared outside the project without proper anonymization. The control over data sharing is rigorous, with established approval processes for sharing any sensitive or personal data. Particularly privacy and appropriate use of AI techniques will be challenged to ensure that proprietary or personal data is not externally accessible and that it avoids over-identification from data streams. Task 1.4 also involves the continual assessment and improvement of knowledge exchange requirements, ensuring research data and outputs are managed according to the FAIR principles, as discussed in section 5. Finally, deliverables D2.2 to D2.4 (which include a detailed description of the business models, functional specifications for interoperability, and standardized data integration) are subject to various aspects of ethical concerns in the context of intelligent energy management systems. These deliverables are designed to address such ethical aspects to guarantee that all the respective dimensions are considered in the architecture of the project's solutions.

Compliance with Policies and Regulations: WeForming also considers various national, sectorial, company, institutional, or departmental procedures and policies related to data management. The project aligns its data management practices with these varying requirements, ensuring compliance across different jurisdictions and organizational contexts. This consideration is particularly relevant in the context of multinational and multi-institutional collaborations within the project. Any specific policies or regulations that impact data management within the project are documented and reported, ensuring full transparency and adherence to all applicable guidelines.

More specifically, T2.3 analyses various issues on the development and adoption of grid-interactive buildings' technologies and solutions, focusing on societal, regulatory, economic, and other types of barriers. It includes an assessment of EU policies and regulatory frameworks and identifies the main regulatory gaps related to distributed energy resources assets and consumer engagement. T4.4 (WeForming Interoperable System Integration) includes the integration and deployment of the project's digital platform, which involves managing data access across various actors in the energy ecosystem and external services/platforms in an interoperable manner, ensuring GDPR compliance through a **Data Protection Impact Assessment (DPIA)**, thereby aligning data management with privacy regulations. Similarly to what has been mentioned about T4.5, this tasks also ensures that the solutions are not only

secure but also that they comply with privacy regulations. It addresses the need for fine-grained access control, anonymization, and encryption across all architectural components, aligning data management with security and privacy standards. Finally, D4.1 includes the description of interoperability enablers within the WeForming framework, detailing the catalogue of services, the data space connector, the middleware, and integrated systems. Since it covers security, privacy, and data sovereignty considerations, it contributes to aligning data management practices with various compliance requirements.

These extended views reflect a comprehensive approach to data management ensuring that all practices are not only technically sound but also ethically responsible and compliant with relevant security, privacy, and regulatory standards. This approach is deeply integrated with the project's specific deliverables and tasks (as each component plays a crucial role in ensuring that data management adheres to various national, sectoral, company, institutional, or departmental procedures and policies, thereby facilitating compliance across different jurisdictions and organizational contexts), enhancing the overall effectiveness and integrity of data management throughout the course of the project.

7. Data Management Process

The Data Management Process (DMP) in WeForming synthesizes all the information provided in previous sections. This process consists of several phases:

Initialize: In this initial phase, the methodology for the DMP is established, and the data management process is initiated. This phase is set in motion by the project's commencement and the definition of key concepts and processes. The groundwork for this phase is laid out in the context of this deliverable, which outlines the data management and IPR protection procedures for the project.

Assess: This phase involves careful monitoring of the data life-cycle stages, utilizing a data management template to record information about the data as it progresses through the life-cycle. The template, defined in section 7.1, also evaluates the adherence to FAIR principles and ethical and security considerations. The assessment is conducted by project partners most familiar with the data, across different pilots where the data originates. Additionally, the assessment aligns with the catalogue entry life-cycle development and any changes or updates to the catalogue entry template. The catalogue is treated as an independent data resource, and its progress is carefully tracked.

Manage: Management of the data management process and influence on the data life-cycle stages are central to this phase. The process aims to be efficient and guides the life-cycle stages, encouraging proper assessment and sharing of data. This includes selecting and recommending potential repositories for publishing data. Management is facilitated through general and technical meetings, as well as dedicated data workshops. Key tasks in this phase include coordinating with Task 2.6 (Functional Specifications for Interoperability and Standardized Data Integration between iGFBs and Their Energy Eco-system) for ensuring quality, quantity, security, and privacy of shared data. Data Protection Impact Assessments are also carried out as part of this task. From a catalogue perspective, this phase is involved in ensuring the collection of appropriate information for each entry, maintaining FAIRness in the catalogue, and monitoring the utilization of entries within the project.

Report: Reporting will occur at scheduled intervals, with the next update related to the DMP information to be triggered (if required) on the occasion of the second version of the Project Management Handbook at Month 18, and the final update (again, if required) at the end of the project in Month 36. These updates will be used to reflect any substantial evolution on the data management practices and the inclusion of new data or changes to existing data.

7.1. Data Management Template

The data management template for WeForming is an elaborated table that extends beyond the initial data items introduced in Section 3 of the plan. This extended table includes additional fields such as a data identifier, change log, sections for FAIR attributes, Extended views attributes, as well as rows for the dataset owner and its current status, as illustrated in the adapted Table 5. Each row of the table represents a distinct data set relevant to the WeForming project. This enhancement facilitates comprehensive tracking and management of data sets, ensuring alignment with project objectives and compliance with data management best practices.

TABLE 5: DATA MANAGEMENT TEMPLATE

Field	Description (Data Set)
Identifier	Unique identifier of the data set, consisting of a combination of an abbreviation and version number (e.g., WF-x.y).
Name	The name of the dataset.
Change Log	A log of changes made to the data set, including the date of each change and a description of what was altered.
Description	Description of the dataset, detailed in a manner similar to that provided in relevant WeForming deliverables (e.g., D2.2, D3.1).
Type	The type of dataset, as categorized in relevant WeForming deliverables (e.g., D2.2 for business models, D3.1 for software building blocks).
FAIR	Conformance to FAIR principles as per the project's guidelines, denoted with letters F (Findable), A (Accessible), I (Interoperable), R (Reusable).
Extended	Status regarding extended views, denoted with: • S (Security measures applied), • P (Privacy aspects considered), • A (Anonymized if necessary), • E (Ethical considerations), • O (Other relevant issues).
Owner	The owner or caretaker of the dataset, typically one of the partners in the WeForming project.
Status	The current status of the dataset, indicating if it has been shared or not, along with the repository where it is stored (if applicable).

Adaptation: The design of this data management template is intentionally straightforward to simplify the monitoring of data sets assessed in the WeForming project. It is understood that not all data sets will be appropriate for sharing; for those that are shared, the template's information will be augmented to comply with the target repository's requirements. Expanded dataset information, along with details about the chosen repository, will be included in an appendix of a subsequent version of this deliverable. This ensures that the data management process is transparent, efficient, and adaptable to the evolving needs of the project.

7.2. Allocation of Resources

The current allocation of resources within the WeForming project for data management is primarily within Work Package 1 (WP1). Specifically, a man-month is allocated for authoring this deliverable. In addition, all project partners have dedicated resources within Task 1.4, which is focused on ethics, exchange requirements specifications, and data management. These resources are earmarked for contributing to

the data management process, which includes preparing deliverables and executing data management tasks as outlined in Section 8 of the Data Management Plan.

Regarding financial considerations, the potential costs associated with open access publishing have been factored into the budgets of some partners. This proactive approach ensures that any publications resulting from the WeForming project can be made freely accessible, enhancing the project's outreach and impact. However, it is important to note that the resources for long-term preservation of data have not been specifically discussed or allocated yet. This aspect will require further consideration and planning to ensure sustainable and effective long-term data preservation strategies are implemented.

8. Data Management Process Implementation Plan

8.1. Data Archiving Implementation

Careful consideration is given to the choice of a platform for archiving and preserving our datasets. Selecting a suitable repository [6] involves assessing several key factors to ensure effective data management and accessibility. These factors include:

- **Assignment of Unique and Persistent Identifiers:** Each dataset should receive a unique and persistent identifier. This is crucial for reliable citations, enabling clear tracking of research outputs to specific researchers and contributions from different grants.
- **Creation of Dedicated Dataset Pages with Rich Metadata:** The repository should create individual pages for each dataset, complete with comprehensive meta-data. This enhances discoverability, aids in understanding the data, facilitates linkage to publications, and supports proper citation, thus promoting the visibility and reuse of research.
- **Tracking of Dataset Usage:** Features for tracking how data is accessed and downloaded are important. These statistics help in understanding the impact and utilization of the data.
- **Alignment with Community Needs and Trustworthiness:** The chosen repository should not only meet the specific needs of the WeForming community but also ideally possess a 'trustworthy data repository' certification. This reflects a commitment to the long-term preservation and accessibility of data.
- **Compliance with Specific Data Requirements:** It's important that the repository aligns with specific data management needs, such as accepted data formats, provisions for access, backup and recovery, and service sustainability. Detailed information about these aspects is usually available on the repository's policy pages.
- **Clear Data Citation Guidelines:** The repository should provide clear instructions on how to appropriately cite any data that is deposited.

8.1.1. PROOFHUB AND DROPBOX

WeForming utilizes ProofHub, a versatile project management software solution developed by ProofHub LLC in 2011 [7], to streamline team collaboration and project coordination, suitable for issue tracking and managing datasets throughout their life-cycle. As a browser-based application, ProofHub offers a suite of tools including a task and deadline calendar, file storage spaces, chat functions, and the ability to manage multiple activity streams through distinct topics. Proofhub will facilitate the backlog management and the tracking of datasets along with the template information outlined in Section 8.2.

In conjunction with ProofHub, WeForming also employs Dropbox, a well-known cloud-based file hosting service. This platform facilitates the storage, synchronization, and sharing of files and folders across various devices, enhancing collaborative efforts among project participants.

As the project coordinator, EUROPEAN DYNAMICS has implemented the use of ProofHub for managing WeForming's internal project data. Additionally, Dropbox is utilized for internal maintenance and backup purposes. Access to WeForming's ProofHub system is meticulously managed by ED, ensuring that only authorized partners within the WeForming consortium have access. This approach underscores our

commitment to efficient project management and secure data handling, aligning with the collaborative and innovative spirit of the WeForming initiative.

8.1.2. ZENODO

Zenodo [8] is a widely recognized research data archive and online repository, facilitating the sharing of research results across various scientific fields. Originating from the European Commission's OpenAIRE+ project [9], Zenodo now operates under the auspices of CERN, leveraging one of Europe's most robust hardware infrastructures. The platform ensures data security through nightly backups and replication across multiple locations.

Zenodo's capabilities extend beyond just the publication of scientific papers and white papers; it also supports the dissemination of structured research data, for example, in XML format. An added feature is Zenodo's GitHub connector, promoting open collaboration on source code and offering version control for diverse data types. All uploaded content is methodically organized using detailed meta-data, such as contributor names, keywords, date, location, document type, license, and more, with English being the preferred language for textual meta-data items. Importantly, all meta-data on Zenodo is licensed under the Creative Commons 'No Rights Reserved' [10] (CC0) license, ensuring wide accessibility. It's crucial to note that uploading results to Zenodo does not alter the ownership or intellectual property rights of the content.

In line with our commitment to open access and data sharing, WeForming will utilize Zenodo for the long-term storage and dissemination of all public results associated with our scientific publications, ensuring they are readily accessible to the wider research community.

8.1.3. CODE REPOSITORY: GITHUB

Within the WeForming project's technical framework, two primary types of repositories are designated for storing the developed programming code. Specific proprietary tools will be stored in private repositories or infrastructures owned by individual project partners. Access to these private repositories will be granted either to all WeForming consortium members or to selected groups specifically involved with the relevant tools.

For the open-source components of WeForming, the technical team is considering various open-source code repositories, with GitHub[11], being a primary option. GitHub is a highly recognized online platform designed to support the development, management, and version control of distributed source code. It is widely used for managing source code data and encourages global collaboration among developers. GitHub also provides features for documentation and issue tracking.

GitHub's service offerings include both paid and free plans. The free plan allows for an unlimited number of public, open-access repositories with an unrestricted number of collaborators, making it an ideal choice for open-source projects seeking broad dissemination. However, private repositories, which are not publicly accessible, require a paid subscription. GitHub is preferred by many open-source initiatives for its ease of sharing results. It organizes projects and their outputs using meta-data like contributor usernames, keywords, timestamps, and file types. According to GitHub's terms of service, the company does not claim any intellectual property rights over the content hosted on its platform. For textual meta-data items, English is the preferred language.

8.1.1. WEFORMING PLATFORM

The WeForming platform is designed as a comprehensive ecosystem for managing and integrating diverse datasets, both structured and unstructured, related to intelligent building energy management. This platform aggregates data from various sources within iGFBs processing it either in batches for services based on aggregate analytics and historical data, or via data streaming technologies for near real-time applications. Post-ingestion, the data undergoes processes for quality enhancement, homogenization, and modelling, thus enabling efficient sharing with users and transformation into formats suitable for data analytics services. This data is then stored for querying and utilization by energy analytics services and platform users.

A key component of the platform's data management services is robust security and access control. This system ensures that only authenticated and authorized users and services can access necessary resources, safeguarding user data stored in relational databases. The security framework, detailed in Section 8.4, provides essential features like data encryption, vulnerability detection and mitigation, and behaviour monitoring and auditing of different entities. As WeForming is in its development stages, the specifics of the platform's functionalities are being actively formulated, laying the groundwork for a secure, efficient, and interoperable data management system within the realm of intelligent energy management in buildings.

8.2. Proposed Workflow for Data Management

ProofHub offers two default workflows: Basic Workflow (that includes two standards work completion stages, To-Do and Done) and KanBan Workflow (which offers three stages; Backlog, In-Process and Complete). The tool also allows for the dynamic management of more workflows and the creation of custom processes. An issue tracking workflow being used since the beginning of the project is comprised of 4 stages, as illustrated in Figure 2:

- **Backlog:** This is where all tasks begin their journey. In this stage, tasks are essentially in a queue, waiting to be actioned. It's a planning and organization phase where tasks are identified and defined but not yet in active progress. This stage is crucial for prioritizing work and ensuring that all necessary tasks are accounted for before the actual work begins.
- **Doing:** Once tasks are moved out of the Backlog, they enter the "Doing" stage. This indicates that work on these tasks has actively started. It's the execution phase where the bulk of the work happens. Tasks in this stage are being worked on by team members, and it's where most of the project's progress is visible. Keeping tasks updated in this stage is vital to provide a clear view of what's currently in progress.
- **Blocked:** This stage is critical for managing impediments and challenges. When a task cannot progress due to an external dependency, lack of resources, or any other blocker, it is moved to the "Blocked" stage. This helps in quickly identifying issues that are hindering progress and need attention. It's essential for efficient project management as it allows teams to address and resolve obstacles promptly.
- **Done:** Tasks that are completed are moved to the "Done" stage. This signifies that the tasks have been accomplished and no further action is required on them. It's a stage for review and closure, providing a clear indication of the project's milestones that have been achieved. It's also useful for retrospective analysis, allowing teams to reflect on completed work, assess the effectiveness of their approaches, and gather insights for future projects.

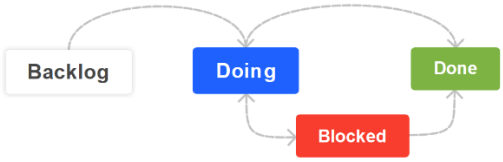

FIGURE 2: CURRENT ISSUE TRACKING PROCESS IMPLEMENTATION

In the context of the WeForming project, this workflow is effectively utilized to manage various tasks, from data collection and processing to ethical reviews and reporting. Each stage of the workflow provides a structured approach to handling different aspects of the project, ensuring clarity, efficiency, and effective progression from task initiation to completion.

As regards the general content / data management process for WeForming, this could be implemented using the following workflow, as illustrated in Figure 3:

- **Open:** This phase involves initializing new datasets within the project. Tasks or items representing these datasets can be created with initial descriptions and relevant meta-data. This stage corresponds to the early planning and first steps in data collection and initial assessment.
- **In Progress:** At this stage, the datasets are actively being worked on. Proofhub can be used to track the ongoing processing, storage, and initial usage of the data. Tasks can be updated with progress notes, file attachments, and discussions among team members. This stage aligns with continuous evaluation and refinement of the datasets.
- **Under Review:** When datasets are ready to be shared or need final approval, they move to the 'Under Review' stage. This can be managed by moving the task to a dedicated list or section for review. Here, datasets are checked for compliance with FAIR principles, ethical considerations, and readiness for external sharing.
- **Published/Rejected:** In this phase, a decision is made on whether the datasets will be published or need to be rejected for specific reasons such as privacy concerns or lack of compliance. Tasks can be tagged as 'Published' or 'Rejected' based on the decision, and notes can be added to provide context for the decision.
- **Report:** The final phase involves documenting and reporting the status of the datasets. This can be handled by generating reports or exporting data that summarize the status and history of each dataset, which can then be included in project reports or updates.

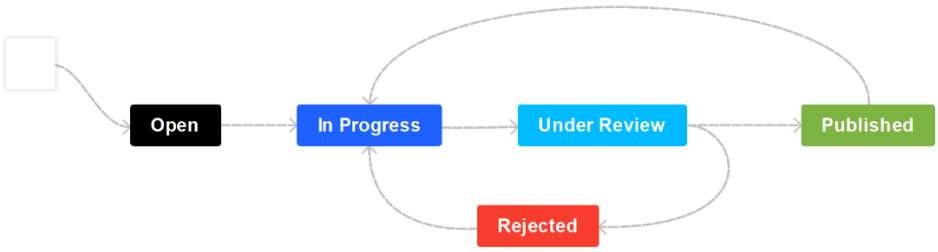

FIGURE 3: POTENTIAL DATA MANAGEMENT PROCESS IMPLEMENTATION

Throughout this process, Proofhub's features such as task lists, notes, discussions, and calendar can be effectively used to manage, track, and document the progress and status of each dataset within the

project. Thus, the tool's features also enable the communication and coordination among team members involved in the data management process.

8.3. Allocation of Resources

In the context of the WeForming project, the costs associated with data storage and maintenance post-project completion are expected to be manageable without requiring additional funding. The value of the data generated may be significant, particularly given the current and evolving needs of the energy sector and customer requirements. While this data is anticipated to have an immediate impact in the coming years, its relevance may evolve as the sector's challenges and priorities shift.

The Project Coordinator (ED) holds the responsibility for maintaining the project's document repository. Concurrently, the Scientific Coordinator (LIST) is tasked with ensuring the quality of all scientific data outcomes. Additionally, each consortium partner in WeForming is responsible for the recoverability of the data they produce. At this stage, the costs for making the data adhere to the FAIR principles are not yet fully quantifiable, as they will largely depend on the total volume of data generated throughout the project's lifespan.

8.4. Data Security

Task 4.5 focuses on establishing robust data security measures through a comprehensive privacy and cybersecurity framework. This framework is designed to ensure the highest levels of security, incorporating detailed access management, anonymization, and encryption for all elements of the platform. To achieve this, the project employs a variety of security methods, including decentralization, authentication, authorization, auditing, policy-based management, and data encryption. Task 4.5 aims to develop a secure and operational platform by integrating these elements, resulting in a prototype ready for validation.

Additionally, the project ensures data security through the application of its established security protocols and frameworks. The data security concept emphasizes the direct involvement of end-users within the 'consumer-centric European data economy,' facilitating mutual benefits for all project participants. The reference architecture, with its focus on interoperability layers, further contributes to a mature, robust, and secure data exchange platform solution.

8.5. Objectives' Ethics, Methodology and Impact

In the WeForming project, various protocols are put in place to ensure the privacy of participating end-users is safeguarded. The consortium overseeing the project will strictly regulate access to information and enforce necessary restrictions. Key measures include:

- Maintaining **adherence to ethical standards and guidelines in line with those of Horizon Europe**, regardless of the location of demonstrations.
- Providing participants with **clear descriptions of the project and study objectives**, presented in an easily understandable manner.
- Highlighting the **voluntary nature of participation** in the project's studies.
- Offering **full disclosure on privacy rights and potential impacts on participants' lives**, along with details about privacy protection measures such as **anonymization** and **secure data storage**.

- Clearly explaining the **time commitment and effort required** for participating in activities.
- Clarifying the **rights of participants to withdraw at any time**, including the option to request the destruction of their personal data.
- **Making available contact information for project stakeholders**.

The consortium will ensure compliance with ethical codes through ongoing reporting processes. In case of human involvement, participants will be thoroughly informed about privacy, confidentiality, and adherence to national and EU legislation. Clear Information Sheets and Informed Consent Forms detailing the voluntary nature of participation, potential risks and benefits, and procedures for incidental findings will be provided.

Participants will have the opportunity to ask questions and receive comprehensible responses. They may also withdraw themselves and their data at any point without adverse consequences. Signed copies of consent forms will be given to the participant or their legal representative, with originals kept in the participant's research record.

Only anonymized or aggregated data, disconnected from individual identification, will be used for workshops and events. Similar data types will be used for dissemination purposes. If personal data processing is essential under certain conditions, the responsible partner will appoint a Data Protection Officer to oversee GDPR compliance and ensure authorized processing of personal data.

If mandated by European and national legislation, relevant authorities must grant authorization for personal data processing. In cases where platforms like Twitter, LinkedIn, Facebook, or Google Cloud are used, which may involve personal data, a Data Processing Addendum/Agreement will be obtained by the responsible partner.

Lastly, WeForming's methodology does not involve clinical trials or children's participation. It also does not anticipate environmental damage, stigmatization of specific social groups, adverse political or financial consequences, misuse, and other similar impacts.

8.6. Compliance

The WeForming consortium is acutely aware of the potential ethical, fundamental rights, privacy, and data protection issues that may arise from the project's activities. As such, it is committed to upholding the highest standards of ethics, fundamental rights, and legal compliance as recognized by the European Union and international bodies. Specifically, the project will align with key ethical principles and fundamental rights as detailed in:

- **The Helsinki Declaration Administrative forms,**
- **The European Code of Conduct for Research Integrity (ECCRI, 2011),**
- **The EU Charter on Fundamental Rights (CFREU, 2010),**
- **The UNESCO Universal Declaration on Bioethics and Human Rights (2005),**
- **The European Convention for the Protection of Human Rights and Fundamental Freedoms (ECHR, 1950),**
- **The Universal Declaration of Human Rights (UDHR, 1948).**

In addressing privacy and personal data protection, WeForming will adhere to:

- **The General Data Protection Regulation (GDPR) (EU) 2016/679,**
- **The Data Protection Directive (1995/46/EC) and the Directive on Privacy and Electronic Communications (2002/58/EC),**
- **The EU Charter on Fundamental Rights (articles 7 and 8),**
- **The European Convention for the Protection of Human Rights and Fundamental Freedoms (article 8),**
- **The CoE Convention No. 108 for the Protection of Individuals concerning Automatic Processing of Personal Data (1981),**
- **The International Covenant on Civil and Political Rights (ICCPR, 1966).**

The Consortium is also prepared for potential changes in the European data protection framework during the project's duration and commits to adhering to any new privacy and data protection regulations. Furthermore, the Consortium will comply with all relevant national and local regulations that apply to the project's activities.

9. Conclusion and Future Work

This deliverable has laid the foundation for the project's data management strategy by establishing a comprehensive framework aligned with FAIR (Findable, Accessible, Interoperable, and Reusable) data management principles. A detailed data life-cycle has been proposed, resonating with the specific needs and objectives, alongside a defined data management process that aligns with the project's diverse and innovative approach.

With this deliverable, the initial stages of the data management process have been activated, focusing primarily on assessment and management. To support this process, a sophisticated project management tool, suitable for the project's needs, is proposed. This tool will assist in implementing an effective data management workflow, simplifying the tracking and assessment of datasets. In the upcoming period, this tool will be set up, and the proposed workflow will be operationalized.

The plan is to introduce this process and tool to the WeForming project partners, using real-world data from the project's pilot studies for demonstration. This practical approach will ensure that all partners are well-versed in the data management practices and can effectively contribute to the project's goals. Future iterations of this DMP will reflect the ongoing evolution and learnings from the project, ensuring that the data management remains dynamic, responsive, and aligned with the project's expanding scope and emerging challenges.

References

[1] European Commission (2024). Continuous reporting on milestones & deliverables - Online Manual. Funding & Tenders Opportunities. Online; Retrieved 12-March-2024 from https://webgate.ec.europa.eu/funding-tenders-opportunities/pages/viewpage.action?pageId=1867968/

[2] FAIR Principles: Interpretations and Implementation Considerations published in "Data Intelligence" (2020; 2 (1-2): 10–29), available at https://doi.org/10.1162/dint_r_00024.

[3] European Commission (2024). Horizon 2020 Online Manual: Cross cutting issues - Open access & Data management. Online; Retrieved 12-March-2024 from https://ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/open-access-data-management/data-management_en.htm

[4] DCMI Usage Board. (2012, June 14). Dublin Core™ Metadata Element Set, Version 1.1: Reference Description. Dublin Core Metadata Initiative. Online; Retrieved 26-March-2024, from https://www.dublincore.org/specifications/dublin-core/dces/

[5] International Data Spaces Association. Components. In IDS Knowledge Base. Online; Retrieved 26-March-2024, from https://docs.internationaldataspaces.org/ids-knowledgebase/v/ids-g/components

[6] European Commission. Open access & Data management - H2020 Online Manual. Online; Retrieved 26-March-2024, from https://ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/open-access-dissemination_en.htm

[7] ProofHub. Manage workflows. ProofHub Help & Support. Online; Retrieved 12-March-2024 from https://help.proofhub.com/plus/tasks/manage-workflows/

[8] .Zenodo org, from https://zenodo.org/

[9] OpenAIRE. Online; Retrieved 26-March-2024, from https://www.openaire.eu/

[10] Creative Commons. CC0. Online; Retrieved [26-March-2024], from https://creativecommons.org/share-your-work/public-domain/cc0/

[11] Wikipedia contributors. (2024, March 21). GitHub. In Wikipedia, The Free Encyclopedia. Online; Retrieved 26-March-2024, from https://en.wikipedia.org/wiki/GitHub

Appendices

9.1. Appendix A: Catalogue entry Template

The WeForming Catalogue Entry Template is designed to describe solutions and datasets within the project. Each item in the template should be briefly described, with references to external documentation and implementations as necessary. These entries are relevant at multiple levels of the WeForming project's focus on iGFBs and energy ecosystems, so the relevance of catalogue entries is defined for various entities, including but not limited to:

- **Energy Efficient Building Assets:**

  This category includes information on various building assets that contribute to the energy efficiency and interactivity of the buildings within the iGFBs. It encompasses a range of technologies and systems used in the building energy ecosystem.

- **Interoperability Solutions:**

  Detailed descriptions of solutions that enhance interoperability within the building energy ecosystem. This might include middleware layers or architectural frameworks that allow seamless integration of building energy management systems with the broader energy network.

- **Energy Management Systems:**

  Systems and tools that are integral to the management of energy in iGFBs, including their interaction with energy markets and networks. This could involve advanced data analytics tools, AI/ML models for predictive energy management, or innovative energy storage solutions.

- **Flexibility Service Providers:**

  Solutions or systems that enable buildings or clusters of buildings to offer flexibility services to the energy grid, contributing to the balance and efficiency of the energy ecosystem.

Catalogue Entry Template:

1. **Name:**

   Unique entry name, potentially an URN at later stages.

2. **Version:**

   Version of the entry description.

3. **Owner:**

   The owner or initiator of the entry/solution, typically a project partner.

4. **Description:**

   A concise description of what the entry provides.

5. **Class:**

   Class of the entry (e.g., Building Management System, Energy Optimization Solution).

6. **Pilot Source:**

   Where the entry will be developed and initially tested (e.g., specific pilot sites).

7. **Use Cases:**

   How the entry relates to use cases and objectives.

8. **Semantic Specification:**

   Any semantic specification used (e.g., industry standards or custom specifications).

9. **Contribution to Standards:**

   Explain how the entry contributes to or aligns with relevant standards and frameworks (e.g., IDSA, EDS).

10. **Interfaces:**

Interaction Models/Interfaces				
IM (Interaction Model)	Sender	Receiver	Attribute[type]	Description
....	...	...	...	...