SEEDS

Cost-effective and replicable RES-integrated electrified heating and cooling systems for improved energy efficiency and demand response.

D1.5 - Data Management Plan (V1)

WP1, Task 1.6

Date

30.06.2024

Author(s): Razgar Ebrahimy (DTU)

DISCLAIMER

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting authority can be held responsible for them.

COPYRIGHT NOTICE

DELIVERABLE INFORMATION

Grant agreement	101138211
Project title	Cost-effective and replicable RES-integrated electrified heating and cooling systems for improved energy efficiency and demand response.
Project acronym	SEEDS
Project coordinator	Danmarks Tekniske Universitet - DTU
Related work package	WP 1
Related task(s)	Task 1.6 – Legal, Knowledge, and Data Management
Lead organisation	DTU
Due date	30/06/2024
Submission date	30/06/2024
Dissemination level	PU

Co-funded by
the European Union

REVISION AND HISTORY CHART

Version	Date	Editors	Description
V0.1	1st March	R2M (Anaïs, Mohi)	First draft for review by the project coordinator
V0.2	12 March	R2M (Anaïs, Mohi)	Full draft ready for peer-review prior to submission

About the Project

The SEEDS project represents a pioneering effort to enhance energy efficiency and promote the electrification of thermal demand in buildings, a crucial step toward decarbonization. It brings together a multidisciplinary consortium of Small and Medium-sized Enterprises (SMEs), Large Enterprises (LEs), Research and Technology Organizations (RTOs), and various stakeholders, encompassing the entire local value chain in the sector of building energy efficiency from across Europe. This encompasses the stages from planning and design, through construction, to operation and commissioning.

Leveraging the consortium's extensive expertise in decarbonization solutions, SEEDS focuses on the development, testing, and application of innovative strategies for building renovation and smartification, along with the deployment of energy flexibility solutions. At the heart of SEEDS are replicable heat pump technologies integrated with renewable energy sources, designed to significantly reduce the carbon footprint of building thermal demand. Recognizing the uniqueness of each building, SEEDS prioritizes the development of scalable, cost-efficient, and energy-efficient solutions tailored to specific needs, thus offering a broad spectrum of optimization methodologies for design and operational efficiency.

SEEDS highlights its groundbreaking solutions through six pilot sites across Europe, including a replication site (Denmark, Slovenia, Belgium, Hungary, and Greece). These sites are strategically selected to represent a wide array of climate zones and construction markets, providing tangible, real-world examples of SEEDS' capabilities up to Technology Readiness Levels (TRL) 6-8.

The initiative is structured around three core themes: enhancing cost efficiency through advanced optimization techniques, achieving system integration via holistic design and control, and ensuring replicability using modular configurations and adaptable building types. To address these themes effectively, SEEDS has identified seven key focus areas, ranging from iterative design processes and secure data management to system optimization, energy flexibility, replication strategies, decision-making support, and comprehensive stakeholder engagement.

With twenty-six partners contributing their diverse expertise and networks, SEEDS is uniquely positioned to make a significant impact on the electrification of thermal demand in buildings.

Deliverable SUMMARY

The Data Management Plan (DMP) is a crucial component of effective data management. The DMP outlines the data management life cycle for the data to be collected, processed, and/or generated by the SEEDS project. To ensure that research data is findable, accessible, interoperable, and reusable (FAIR), the DMP includes information on:

- Types of data, specifying what data will be collected, processed, and/or generated.
- Data collection and methods, detailing the methodologies and standards that will be applied.
- Data sharing and transparency, including information about open access to data.
- Data storage and archiving, explaining how data will be curated and preserved (including after the project ends).
- Intellectual rights, licenses, data security, and ethics.

This deliverable will be update internally in Month 18, in Month 36 and in Month 48 of the SEEDS project.

TABLE OF CONTENTS

Co-funded by
the European Union

LIST OF FIGURES

No table of figures entries found.

LIST OF TABLES

Co-funded by
the European Union

Table 1 List of Acronyms

Acronym	Description
DMP	Data Management Plan
FAIR	Findable, Accessible, Interoperable, and Reusable
WP	Work Package
WPLs	Work Package Leaders
PLs	Pilot Leader
GA	Grant Agreement
IPR	Intellectual Property Rights
GDPR	General Data Protection Regulation
ENISA	European Union Agency for Network and Information Security guidance
EC	European Commission
ORD Pilot	open access to research data
RRI	Responsible Research and Innovation

1.INTRODUCTION

1.1. AIMS AND OBJECTIVES

This deliverable contains five main chapters that explain the data management process as follows:

- Description of data, formats, organizing, metadata (Chapter 2)
- Data collection and methods (Chapter 3)
- Data sharing and transparency (Chapter 4)
- Data Storing and archiving (Chapter 5)
- Intellectual rights, licenses, data security, and ethics (Chapter 6)

The data management plan (DMP) will ensure regular discussions with all partners and will be updated internally to maintain consistent, accurate, and reliable data analysis throughout and beyond the four-year life cycle of the SEEDS project. Additionally, some data will be monitored, analyzed, and stored for extended periods, provided data owners permit it.

Data generated in the project will adhere to the principles of being 'Findable, Accessible, Interoperable, and Reusable (FAIR),' in accordance with the H2020 Guidelines on FAIR Data Management and the Guidelines on Implementation of Open Access to Scientific Publications and Research Data (European Research Council, 2017).

This is the first version of the DMP, outlining the overall data management approach in the project. It includes the lifecycle of data to be collected, generated, used, or processed within the project, and addresses data handling methodologies, data sharing, privacy and security considerations, legal and regulatory requirements, informed consent, and open access during and after the project. This deliverable is part of Task 1.6 under Work Package (WP) 1, Project Management and Coordination.

The DMP will be further detailed and structured as an evolving document, following the overall data governance from WP1, with additional input from all other WPs and Pilot Leaders (PLs) as appropriate.

The purpose of this report is to provide a comprehensive overview of all data anticipated to be produced, used, and re-used in the SEEDS project. It next version will detail the types of data expected to be generated, the strategies for making this data open and accessible for verification and re-use, and the methods for curation and preservation. Additionally, it addresses critical aspects related to intellectual rights, licenses, data security, and ethics.

This report offers a preliminary overview of data handling within the project, setting forth the initial guidelines for managing data throughout the project's lifecycle. It aims to ensure that data is handled efficiently, securely, and ethically, promoting transparency and facilitating future research and application. These guidelines will serve as a foundational framework, aiding in consistent data management practices across all work packages and the pilot implementation of the SEEDS project.

2. DESCRIPTION OF DATA, FORMATS, ORGANIZING, METADATA

This chapter provides an overview of the anticipated SEEDS data, organized by WPLs and PLs. It includes the following information:

- Origin and Source of Data: Details the origin and source of the data.
- Reuse of Existing Dat: Identifies any existing data to be reused and the licensing terms and conditions (Are the data publicly available? Can the data be used, modified, or distributed?).
- Expected Data Size: Provides an estimate of the data size.
- Responsibility: Specifies who is responsible for data collection or development.

In further development of the DMP, the following topics will also be included:

- **Data Set Reference and Description**: Identifier for the data set to be produced and a description of the data set.
- **Data Description**: Description of the data to be generated or collected, its origin (if collected), nature, scale, potential usefulness, and whether it underpins a scientific publication. It also includes information on the existence of similar data and the possibilities for integration and reuse.
- **Standards and Metadata**: Details on data standards and metadata, and data sharing protocols.
- **Reference Standards**: Reference to existing suitable standards of discipline. If these do not exist, an outline of how and what metadata will be created.

The consortium will provide Open Data arising from the project to support the following goals:

1. **Benchmarking**: Comparing developed measures with other projects.
2. **Dissemination and Exploitation**: Enhancing the dissemination and exploitation of data and results.

3. **Access and Reuse**: Improving access to and reuse of research data generated within the project.
4. **Knowledge Sharing**: Ensuring knowledge-sharing with citizens, the wider public, interested stakeholders, cities, industry, and the scientific community.

A simple format to describe and package a single data resource will be prepared for easier data usage, modification, and distribution. To achieve this, standards such as the Frictionless Standards will be utilized by the consortium. This will enable data discussion and further contribution among partners. Over time, clear definitions will be established to facilitate interactions between different stakeholders inside and outside the consortium.

As of writing this deliverable, there are four main categories of data arising from the project:

1. **Underlying Research Data**: Data necessary for validation of results presented in scientific papers, including associated metadata, which aligns with the general principle of openness of scientific results.
2. **Operational and Observational Data**: Includes curated or raw data from the implementation, testing, and operation of demonstrators (operational data), as well as data from related qualitative activities, such as surveys, interviews, fieldwork data, engagement activities, workshops, and personal data (observational data).
3. **Evaluation Data**: Data specifically captured to track KPIs of the project performance, regularly reported and published.
4. **Documentation, Instruments, and Reusable Knowledge**: Includes general and specific documentation of the project and demonstration/implementation projects, tools, methods, instruments, software, and underlying source code needed to replicate the results.

Table 2 Data roles and responsibilities example

Role	Name	Data Types	Data Origin	Data size	Accessibility and data sharing	Preservation and storing	Data re-using Data utility	Responsibility
WP1	DTU	Contact list	Partners	0.5Mb<	Available to all partners	Secure Share Drive	To be used for internal communications	All

3. DATA COLLECTION AND METHODS

The SEEDS project will generate data from both the design and construction/retrofitting phases, encompassing two spatial levels: buildings and neighborhoods/urban areas. Some of this data will be open to the public, while other data will remain the property of the party that collected, produced, or assessed it and only to be shared with the partners involved in the study or pilot site.

The assessment framework will be developed in Task 1.6 will serve as the foundation for common data collection and evaluation methods. The method will be further detailed in the next iteration of DMP deliverables.

The framework will outline the static and dynamic data collection and monitoring processes for the SEEDS demo projects, comprising four steps:

1. **Monitoring Systems Definition**:
    - ❖ Define and describe the monitoring systems in each SEEDS pilot, including a list of technologies, data requirements, and sources.
    - ❖ Cover the physical architectures of monitoring systems, their network architectures, sensors to be installed, and the functioning of data collection processes (frequency of measurements, local storage, backup systems).
    - ❖ Each pilot's monitoring system will be customized, considering inputs from demo lead partners. Monitoring aspects include:
        i. Energy production (on-site monitoring)
        ii. Energy distribution (on-site monitoring)
        iii. Energy consumption (on-site monitoring)
        iv. Indoor comfort
        v. User behavior
        vi. External micro-climate conditions (at district scale)
2. Data Architecture Design:
    - ❖ Follow design data architecture using open standards.
    - ❖ Select a database of technologies to be used.
    - ❖ Define common protocols for data collection with pilot leaders.
    - ❖ Unify data from each monitoring system into a unique database.
3. Data Quality Assessment:
    - ❖ Implement regular data quality checks and provide feedback to respective SEEDS pilots on the quality of their data.
4. Data Warehouse Implementation and Population:
    - ❖ Include data on construction activities and occupancy in new buildings; retrofit activities and pre- and post-intervention occupancy in retrofitted buildings; applied surveys; and simulations.

The comprehensive approach ensures that data is systematically collected, managed, and utilized, supporting the project's goals of transparency, reliability, and accessibility.

Table 3 A general overview of data collection methods in SEEDS project

WP/Pilot	Data Type	Data collection methods and tools
WP1	Observational	Monitoring, meetings, and workshops
WP2	Research/Operational Evaluation/Documentation	Simulations, surveys, VR/AR, reviews, Georeferenced data, Blueprints and 3D Models/Digital Twins, workshops
WP3	Research/Operational Evaluation/Documentation	Monitoring, simulations, fieldwork, testing
WP4	Research/Operational Evaluation/Documentation	Monitoring, simulations, fieldwork, testing
WP5	Research/Operational Evaluation/Documentation	Monitoring, simulations, fieldwork, testing
WP6	Research/Operational Evaluation/Documentation	Surveys, workshops, monitoring, co-creation
WP7	Research/Operational Evaluation/Documentation	Monitoring, simulations, fieldwork, testing, reviews
WP8	Research/Operational Evaluation/Documentation	Social media monitoring (statistics on website and social media performance), surveys, meetings, workshops
Five Pilot sites	Research/Operational Evaluation/Documentation	All data collection methods above

DTU, as the lead partner, is responsible for the overall data management of the SEEDS project. Work package leaders are responsible for data management within their respective WPs, and pilot leaders are responsible for data management within their specific demonstrations.

4. DATA SHARING AND TRANSPARENCY

The project will adhere to the EU principle "as open as possible, as closed as necessary." The SEEDS consortium will encourage the publication of scientific papers, provided it does not impede potential patenting, or the protection of intellectual property rights (IPR) generated within the project. Therefore, prior permission to publish any project-related information must be submitted to the Executive Board (EB) to ensure that sensitive material is not disclosed. If a dataset cannot be shared, the reasons must be explained, such as ethical considerations, personal data regulations, intellectual property, commercial interests, privacy, or security concerns.

The Model Grant Agreement (GA) outlines the detailed legal requirements for open access to scientific publications. Under Horizon 2020, each beneficiary must ensure open access to all peer-reviewed scientific publications. To meet this requirement, beneficiaries must, at a minimum, ensure that any scientific peer-reviewed publications can be read online, downloaded, and printed. Since additional rights—such as the rights to copy, distribute, search, link, crawl, and mine—enhance the utility of publications, beneficiaries should make every effort to provide as many of these options as possible.

5. DATA STORING AND ARCHIVING

Project documentation will be stored and exchanged on a private, secure, and confidential SharePoint platform administered by the Project Coordinator, DTU. All project members will have registered access for internal document exchange. This platform facilitates efficient knowledge management, ensuring data and information exchanges between project members without unnecessary risk of external disclosure.

The elaborative version of Table 1 this report will be used to present detailed information about data accessibility and sharing, preservation and storage of data, and data re-using and redistributing, along with clear roles and responsibilities of all partners. This data will be provided by the responsible WP leaders, who oversee data management for their respective WPs and Pilots, through the established internal data management approach and defined rules, such as:

- For confidential/internal data: managed through a private repository (e.g., GitLab or the control models developed in WP4 and WP5).

- For public data: using GitLab or the data architectue developed by WP3, and linking the datasets with platforms such as Zenodo, EnerMaps, and OpenAIRE to enhance visibility and comply with FAIR requirements.

The aim is to harmonize and discuss a common pathway, including access procedures, embargo periods (if any), outlines of technical mechanisms for dissemination, necessary software and other tools for enabling re-use, and definitions of whether access will be widely open or restricted to specific groups. If needed, the table will be updated annually.

More detailed requirements and documentation will be generated before the start of any activity involving the participation of individuals being subjects of the study. This will be detailed and tailored to all demonstration projects, in the official language of the country/city where the demonstration takes place. It will include demonstration-specific aspects and refer to relevant regulations on data protection and/or other applicable legislation.

As part of WP3 involved will facilitate the establishment of a data lake setup for software, data management, and storage. This setup will include a federated data lake that integrates data analytics from fog, edge, and cloud computing environments, ensuring compliance with

General Data Protection Regulation (GDPR), cyber security, and data ethics. SEEDS will deploy a decentralized digitalization cloud level connected to fog level at individual pilot sites, which in turn are connected to building and occupant systems or apps (edge level). Utilizing a state-of-the-art data lake principle, the SEEDS project will provide storage for monitoring data and enable efficient data use across various scales at each pilot site. It will also offer user interfaces for occupants, operators, building owners, and other stakeholders if needed.

Engineering best practices and state-of-the-art data security measures will be incorporated through a security architecture, GDPR considerations, and respective guidelines and principles. Information security management will follow the guidelines of relevant standards (e.g., ISO/IEC 27001 and 27002 Code of practice for information security management) to ensure confidentiality, integrity, and availability. It may also include the Directive on security of network and information systems (NIS-Directive) 2016/1148 on the security of critical infrastructures and the Privacy Directive 2002/58, along with European Union Agency for Network and Information Security (ENISA) guidance. The data storage will fully comply with national and EU legal and regulatory requirements. Partners will ensure that the cloud infrastructure used complies with applicable regulations.

Furthermore, the SEEDS partner organizations have their own data protection routines established in their existing operations and in their development and test activities within the project, in compliance with GDPR and other data protection and security regulations.

6. INTELLECTUAL RIGHTS, LICENSES, DATA SECURITY, AND ETHICS

This chapter includes a description of legal and intellectual rights, along with other ethical requirements for the SEEDS data.

The European Commission (EC) has launched a pilot for open access to research data (ORD pilot) to improve and maximize access to and re-use of research data generated by Horizon 2020 projects across all thematic areas. This initiative considers the need to balance openness and protection of scientific information, commercialization and intellectual property rights (IPR), privacy concerns, security, and data management and preservation questions. The SEEDS DMP will address all these aspects comprehensively.

All activities within the SEEDS project will be conducted in compliance with fundamental ethical principles and will adhere to the principles and practices of Responsible Research and Innovation (RRI). Project management will take the necessary actions to ensure all beneficiaries comply with applicable European and national regulations and professional codes of conduct related to personal data protection. These include Directive 95/46/EC regarding data collection and processing, the GDPR, 2016/679 that took effect in May 2018, and respective national requirements, ensuring legal and regulatory compliance.

When personal data collection or processing begins, the DMP information will be updated to include new data summaries, consent forms, compliance measures, and institutional approvals where necessary. The processing of personal data will respect Data Protection Principles, as per H2020 Ethics and FAIR guidelines.

Any data relating to the pilot sites, such as utility bills, metered data and sensors data, will remain strictly the property of the pilot sites and will only be shared with the permission of the pilot site owner.

The project will respect the privacy of all stakeholders and users, seeking free and fully informed consent whenever personally identifiable data is collected and processed. Suitable data handling procedures and protocols will be implemented to prevent the potential identification of individuals. This includes data collected from activities such as questionnaires, interviews, workshops, mailing lists, and automatic building and energy data collection. Where necessary, data will be anonymized, and the mapping of IDs to individuals will be safeguarded, accessible only to those directly working with the data. All beneficiaries will handle all material with strict confidentiality and privacy, in accordance with legal and regulatory requirements, ensuring no harm comes to participants, stakeholders, or any unknown third parties.

Results will be used in anonymized or aggregate forms of analysis and may subsequently be published in project reports and scientific papers. The consortium confirms that the ethical standards and guidelines of Horizon 2020 will be rigorously applied, regardless of the country in which the work is carried out, and that all data transfers will comply with all necessary legal and regulatory requirements. The project aims to make anonymized datasets public while balancing data publication with privacy and confidentiality concerns.

7.CONCLUSION

In conclusion, the SEEDS project is committed to exemplary data management practices that prioritize transparency, security, and ethical considerations. Through a comprehensive DMP, the project outlines the lifecycle of data to be collected, processed, and generated, ensuring it is findable, accessible, interoperable, and reusable. By leveraging advanced technologies and adhering to the highest standards, SEEDS will effectively manage data from both the design and construction phases, as well as from building and urban levels.

The project emphasizes the importance of privacy and informed consent, implementing rigorous data handling protocols to protect personally identifiable information. Data will be anonymized where necessary, and access to sensitive information will be strictly controlled. The SEEDS consortium will utilize secure platforms for data storage and sharing, ensuring compliance with European and national regulations, including GDPR.

Open access to scientific publications and research data is encouraged, balancing the need for openness with the protection of intellectual property rights and privacy concerns. The project will adhere to the ethical standards and guidelines of Horizon 2020, ensuring that all data transfers and activities comply with relevant legal requirements.

By fostering a culture of responsible research and innovation, the SEEDS project aims to contribute valuable insights and data to the scientific community, stakeholders, and the public, while maintaining the highest standards of data integrity and ethical responsibility. The collaborative efforts of all partners will ensure the successful implementation of data management strategies, supporting the overall goals of the SEEDS project.

SEEDS

https://project-seeds.eu/

project-seeds

@EU_SEEDS