



# SEEDS

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

## D1.5 – DATA MANAGEMENT PLAN

### WP1, Task 1.6

**Authors:** Rongling Li (DTU), Desi Assentoft (DTU), Naja Kastrup Friis (DTU), Ann Bruggeman (SWECO), Filip Jorissen (BWS), Lone Meertens (KUL), Louis Hermans (KUL), Lieve Helsen (KUL), Mikkel Bendixen Copeland (CDK), Adam S e Beilin (Enfor), Jonas Wiendl (Alnergy), Achilleas Andronikos (CERTH), Bernard  zenko (JSI), Erik Gramc (Petrol), Paolo TRAVAGLI (Rina), Natan Vandeginste (SWECO), Anais Gandelin (R2M)

Date of delivery: 11/02/2026



Co-funded by  
the European Union

This project is co-funded by the European Union's Horizon Europe  
Innovation actions programme under the Grant Agreement n 101138211



## ACKNOWLEDGEMENTS AND 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

©2024 SEEDS Consortium Partners. All rights reserved. SEEDS project is co-funded by the European Union's Horizon Europe innovation actions programme under the Grant Agreement n°101138211. For more information on the project, its partners and contributors, please see the SEEDS website (<https://project-seeds.eu/>). You are permitted to copy and distribute verbatim copies of this document, containing this copyright notice, but modifying this document is not allowed. All contents are reserved by default and may not be disclosed to third parties without the written consent of the SEEDS partners, except as mandated by the European Commission contract, for reviewing and dissemination purposes. All trademarks and other rights on third party products mentioned in this document are acknowledged and owned by the respective holders. The information contained in this document represents the views of SEEDS members as of the date they are published. The SEEDS consortium does not guarantee that any information contained herein is e-free, or up-to-date, nor makes warranties, express, implied, or statutory, by publishing this document.

## PROJECT 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	The Technical University of Denmark (DTU)
Lead organisation	DTU
Website	<a href="https://project-seeds.eu/">https://project-seeds.eu/</a>

## DOCUMENT INFORMATION

Deliverable number and title	D1.5 – Data Management Plan
Due date	30/06/2024 (1 <sup>st</sup> version)
Work package number and title	WP 1 – Project Management
Related task(s) number and title	Task 1.6 - Legal, Knowledge, and Data Management
Dissemination level	PU
Submission date	11/02/2026 (updated version)
Lead author	Rongling Li (DTU), Desi Assentoft (DTU), Naja Kastrup Friis (DTU), Ann Bruggeman (SWECO), Lieve Helsen (KUL), Mikkel Bendixen Copeland (CDK), Achilleas Andronikos (CERTH), Bernard Ženko (JSI), Erik Gramc (Petrol), Paolo TRAVAGLI (Rina), Natan Vandeginste (SWECO), Anais Gandelin (R2M)

---

Contributors	Filip Jorissen (BWS), Lone Meertens (KUL), Louis Hermans (KUL), Adam S�e Beilin (Enfor), Jonas Wiendl (Alnergy)
Reviewers	Marcel Sol� Avila (DTU), Nikos Rapkos (RENEL)

## REVISION AND HISTORY CHART

*Table 1, Presenting the timeline of contributions to this report*

Date	Editors	Description
24.11.2025	Rongling Li, Desi Assentoft, Naja Kastrup Friis	Table of content, First draft of DK pilot, Revision of content according to the FAIR principles- EU Horizon project requirements
12.01.2026	Ann Bruggeman	First draft BE Pilot / Replica
13.01.2026	Filip Jorissen	Providing Builtwins contributions
13.01.2026	Naja Friis	Adjust layout, First draft WP1, review of report
13.01.2026	Rongling Li, Mikkel Bendixen Copeland, Adam Søe Beilin, Jonas Wiendl	Final draft of DK pilot
15.01.2026	Achilleas Andronikos	El pilot input
19.01.2026	Lone Meertens, Louis Hermans	Providing KU Leuven contributions
20.01.2026	Lieve Helsen	Revision general part and BE Pilot
20.01.2026	Bernard Ženko, Erik Gramc	SI pilot input
20.01.2026	Naja Kastrup Friis	WP1 input
22.01.2026	Paolo TRAVAGLI	WP6 input
22.01.2026	Natan Vandeginste	WP7 input
22.01.2026	Anais Gandelin	WP8 input
23.01.2026	Rongling Li	Revision and invite all partners to review
11.02.2026	Rongling Li	Finalization and 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 Organisations (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. Recognising the uniqueness of each building, SEEDS prioritises the development of scalable, cost-efficient, and energy-efficient solutions tailored to specific needs, thus offering a broad spectrum of optimisation 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 optimisation 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 optimisation, 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.

## EXECUTIVE SUMMARY

The objective of this report is to inform how data is collected, used, stored, and preserved during the SEEDS project. The need and relevance of the data are also justified. Compliance with the FAIR principles (Findable, Accessible, Interoperable, and Reusable) is presented for all data, as SEEDS is committed to implementing these throughout all data management activities. The aim of the FAIR principles is to ensure that the output can be verified, built upon and translated into real-world solutions. The report is structured with an introduction to the FAIR principles, followed by presentations of data management for each pilot and the complementary WPs. The types of data are very diverse, including contact information, methodologies for data management and collection, data from simulations and optimisations, and measured data from components and buildings.

# CONTENTS

1. INTRODUCTION .....	11
1.1. FAIR PRINCIPLES .....	11
1.2. REPORT STRUCTURE.....	11
2. COMMON FRAMEWORK APPLICABLE TO ALL PARTNERS – FAIR CRITERIA .	12
2.1. COMMON FRAMEWORK.....	12
2.1.1. Purpose of the data .....	12
2.1.2. Collected data.....	12
Underlying Research Data .....	13
Operational and Observational Data.....	13
2.1.3. Evaluation of Data .....	13
2.1.4. Documentation and Reusable Knowledge .....	13
2.1.5. How Data is used .....	13
During the Project.....	13
After the Project .....	14
2.1.6. Data Collection Framework.....	14
2.1.7. How data is stored .....	14
Internal/Confidential Data .....	14
Public Data and Long-term Preservation .....	14
Security Standards.....	15
2.1.8. Timeline of data collection and preservation .....	15
Timeline for Data Openness .....	15
2.2. FAIR PRINCIPLES IMPLEMENTATION .....	15
2.2.1. Making data Findable .....	15
Persistent Identifiers.....	16
Rich Metadata .....	16
2.2.2. Making data Accessible.....	16
Publication-linked data .....	16
2.2.3. Metadata Accessibility .....	17
2.2.4. Making data Interoperable .....	17
2.2.5. Making data Reusable .....	17
Comprehensive Documentation.....	17
Data Licenses.....	17

Quality Assurance.....	17
2.3.    IMPLEMENTATION ACROSS PARTNERS.....	18
<b>3. DATA MANAGEMENT BY PILOT .....</b>	<b>19</b>
3.1.    DANISH PILOT .....	19
3.1.1.    Purpose of the data.....	19
3.1.2.    Collected data.....	19
3.1.3.    How data is used.....	20
3.1.4.    How Data Is Stored.....	21
3.1.5.    Timeline of data collection and preservation .....	21
3.1.6.    FAIR Criteria .....	22
3.2.    SLOVENIAN PILOT .....	23
3.2.1.    Purpose of the Data.....	23
3.2.2.    Collected data.....	23
3.2.3.    How Data is used .....	24
3.2.4.    How data is stored .....	24
3.2.5.    Timeline of data collection and preservation .....	24
3.2.6.    FAIR Criteria .....	24
3.3.    BELGIUM PILOTS.....	25
3.3.1.    Purpose of the Data.....	25
3.3.2.    Collected data.....	25
3.3.3.    How Data is used .....	27
3.3.4.    How data is stored .....	28
3.3.5.    Timeline of data collection and preservation .....	29
3.3.6.    FAIR Criteria .....	29
3.4.    GREEK PILOT .....	30
3.4.1.    Purpose of the Data.....	30
3.4.2.    Collected data.....	31
3.4.3.    How Data is used .....	31
3.4.4.    How data is stored .....	33
3.4.5.    Timeline of data collection and preservation .....	33
3.4.6.    FAIR Criteria .....	33
<b>4. DATA MANAGEMENT BY WORK PACKAGE.....</b>	<b>35</b>
4.1.    WP1 DATA MANAGEMENT .....	35

4.1.1.	Purpose of the Data.....	35
4.1.2.	Collected data.....	35
4.1.3.	How Data is used .....	36
4.1.4.	How data is stored .....	36
4.1.5.	Timeline of data collection and preservation .....	36
4.1.6.	FAIR Criteria .....	36
4.2.	WP6 DATA MANAGEMENT .....	37
4.2.1.	Purpose of the Data.....	37
4.2.2.	Collected data.....	37
4.2.3.	How Data is used .....	38
4.2.4.	How data is stored .....	39
4.2.5.	Timeline of data collection and preservation .....	39
4.2.6.	FAIR Criteria .....	39
4.3.	WP7 DATA MANAGEMENT .....	40
4.3.1.	Purpose of the Data.....	40
4.3.2.	Collected data.....	40
4.3.3.	How Data is used .....	40
4.3.4.	How data is stored .....	40
4.3.5.	Timeline of data collection and preservation .....	40
4.3.6.	FAIR Criteria .....	41
4.4.	WP8 DATA MANAGEMENT .....	41
4.4.1.	Purpose of the Data.....	41
4.4.2.	Collected data.....	41
4.4.3.	How Data is used .....	41
4.4.4.	How data is stored .....	41
4.4.5.	Timeline of data collection and preservation .....	42
4.4.6.	FAIR Criteria .....	42
5.	CONCLUSION .....	43

## LIST OF TABLES

Table 1, Presenting the timeline of contributions to this report .....	4
Table 2, A general overview of data collection methods in SEEDS project .....	12
Table 3, Technologies used for data collection at different system levels .....	24

## LIST OF FIGURES

Figure 1, Visualisation of end-to-end data flow .....	19
Figure 2, Example of district heating metering data from the Nordomatic CTS system.....	20
Figure 3, Example of dataset management.....	22
Figure 4, Visual Analytics Engine for facility managers in EL pilot .....	32
Figure 5, Notifications and Alerts for facility managers in EL Pilot .....	32
Figure 6, Data Management feature for facility managers in EI Pilot .....	34

## LIST OF ABBREVIATIONS AND ACRONYMS

---

<b>Abbreviation</b>	<b>Description</b>
API	Application Programming Interface
BACnet	Building Automation and Control Networks
BMS	Building Management System
CTS	Control and Technical Systems
DMP	Data Management Plan
DOI	Digital Object Identifier
DTS	Distributed Temperature sensor System
ENISA	European Union Agency for Network and Information Security
FAIR	Findable, Accessible, Interoperable, and Reusable
GDPR	General Data Protection Regulation
HVAC	Heating, Ventilation, and Air Conditioning
IOCS	Integrated Optimal Control and Sizing
IoT	Internet of Things
KPI	Key Performance Indicator
LEs	Large Enterprises
ManGO	Management van Gegevens voor Onderzoek
MPC	Model Predictive Control
NIS-directive	Network and Information Systems
RTOs	Research and Technology Organisations
SMEs	Small and Medium-sized Enterprises
TRL	Technology Readiness Levels

---

# 1. INTRODUCTION

---

This Data Management Plan (DMP) outlines the comprehensive framework for managing data throughout the SEEDS project lifecycle. It ensures that all data generated, collected, processed, and shared comply with Horizon Europe requirements for Research Data Management.

The SEEDS project implies the FAIR data principles (Findable, Accessible, Interoperable, and Reusable) across all consortium partners. This framework ensures compliance with EU Horizon Europe requirements for open science while protecting intellectual property throughout the project period.

SEEDS is structured around the pilot sites with technical work packages (WPs 2-5) developing innovative solutions and other work packages (WPs 1, 6-8) supporting technical development and bringing the innovations to the next level. Therefore, this DMP is structured to cover all types of data:

- Data collected from pilot sites for use in WPs 2-5, and
- Data collected within the project for use in WPs 1, 6-8

This report only provides information on data management. Further details on how the data is used can be found in various deliverables.

## 1.1. FAIR PRINCIPLES

The SEEDS project is committed to implementing the FAIR data principles throughout all data management activities:

- Findable: Data will be assigned persistent identifiers and described with rich metadata to enable discovery through catalogues and search engines.
- Accessible: Data and metadata will be stored in trusted repositories with clear access procedures and protocols, ensuring long-term preservation.
- Interoperable: Data will use standardised formats, vocabularies, and ontologies to enable integration and reuse across different systems and communities.
- Reusable: Data will be released with clear usage licenses, provenance information, and will meet domain-relevant community standards.

## 1.2. REPORT STRUCTURE

This DMP is organised to provide clear guidance on data management for each pilot site and work package:

- Section 2: Common framework applicable to all partners – FAIR Criteria; Current Status; Implementation Across Partners
- Section 3: Data management approaches for each Pilot
- Section 4: Data management approaches for the complementary Work Packages

## 2. COMMON FRAMEWORK APPLICABLE TO ALL PARTNERS – FAIR CRITERIA

---

### 2.1. COMMON FRAMEWORK

This framework establishes common data management practices for all SEEDS consortium partners. It ensures compliance with EU Horizon Europe requirements for open science while respecting intellectual property protection during the project period. The project adheres to the EU principle "as open as possible, as closed as necessary."

#### 2.1.1. PURPOSE OF THE DATA

The SEEDS project collects and generates research data to achieve its core objectives:

- Validate and demonstrate innovative renewable energy solutions across European pilot sites
- Monitor energy performance of buildings and systems to optimise heat pump and storage technologies
- Enable evidence-based policy recommendations for sustainable energy transition
- Support replication of successful solutions in other contexts across Europe
- Contribute to scientific knowledge through peer-reviewed publications

Open Data arising from the project supports benchmarking with other projects, enhanced dissemination and exploitation, improved access and reuse of research data, and knowledge sharing with citizens, stakeholders, cities, industry, and the scientific community.

#### 2.1.2. COLLECTED DATA

---

*Table 2, A general overview of data collection methods in SEEDS project*

WP/Pilot	Data Type	Data collection methods and tools
<b>WP1</b>	Observational	Monitoring, meetings, and workshops
<b>WP2</b>	Research/Operational Evaluation/Documentation	Surveys, reviews, design, simulations, blueprints and 3D Models/Digital Twins, workshops, fieldwork, testing
<b>WP3</b>	Research/Operational Evaluation/Documentation	Surveys, Monitoring, simulations, fieldwork, testing
<b>WP4</b>	Research/Operational Evaluation/Documentation	Monitoring, simulations, optimisations, fieldwork, testing
<b>WP5</b>	Research/Operational Evaluation/Documentation	Monitoring, simulations, optimisations, fieldwork, testing
<b>WP6</b>	Research/Operational Evaluation/Documentation	Surveys, workshops, monitoring, co-creation
<b>WP7</b>	Research/Operational Evaluation/Documentation	Interviews

---

<b>WP8</b>	Research/Operational Evaluation/Documentation	Social media monitoring (statistics on website and social media performance), surveys, meetings, and workshops
<b>Five Pilot sites (primarily for tasks in WPs 2-5)</b>	Research/Operational Evaluation/Documentation	All data collection methods above

SEEDS partners collect and generate four main categories of data:

### UNDERLYING RESEARCH DATA

Data necessary for validation of results presented in scientific papers, including associated metadata. This aligns with the general principle of openness of scientific results.

### 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, engagement activities, workshops, and personal data (observational data).

Monitoring aspects covered across pilot sites include:

- Energy production (on-site monitoring)
- Energy distribution (on-site monitoring)
- Energy consumption (on-site monitoring)
- Indoor comfort measurements
- User behaviour patterns
- External micro-climate conditions (at district scale)

#### 2.1.3. EVALUATION OF DATA

Data specifically captured to track project Key Performance Indicators (KPIs) of the project performance, regularly reported and published.

#### 2.1.4. DOCUMENTATION AND REUSABLE KNOWLEDGE

Includes general and specific documentation of the project and demonstration/implementation projects, tools, methods, instruments, software, and underlying (potentially proprietary) source code needed to replicate the results.

#### 2.1.5. HOW DATA IS USED

Research data collected within SEEDS serves multiple purposes throughout the project lifecycle.

### DURING THE PROJECT

- Performance analysis and system optimisation at pilot sites

- Validation of models and simulation tools
- Development of best practices and guidelines
- Scientific publications and conference presentations
- Policy recommendations to EU and national bodies
- Outreach to practitioners

#### AFTER THE PROJECT

- Enabling replication of solutions by other researchers and practitioners
- Supporting future research and innovation projects
- Benchmarking for industry and policymakers
- Educational and training purposes

#### 2.1.6. DATA COLLECTION FRAMEWORK

The assessment framework developed in Task 1.6 provides the foundation for common data collection and evaluation methods across all pilots, comprising four steps:

- **Monitoring Systems Definition:** Define and describe monitoring systems in each pilot, including technologies, data requirements, sources, physical and network architectures, sensors, and data collection processes.
- **Data Architecture Design:** Follow open standards, select database technologies, define common protocols, and unify data from each monitoring system into a unique database.
- **Data Quality Assessment:** Implement regular data quality checks and provide feedback to pilots on data quality.

#### 2.1.7. HOW DATA IS STORED

Project documentation is 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.

Pilot site data are stored by dedicated partners of each site, e.g., for the Danish pilot, CDK manages and stores the data collected from the pilot site through the established internal data management approach and defined rules specified in section 3.

#### INTERNAL/CONFIDENTIAL DATA

- Confidential SharePoint platform administered by DTU for secure internal document exchange
- Private GitLab repository for confidential code and control models (WP2, WP4, WP5)
- Access restricted to registered consortium members

#### PUBLIC DATA AND LONG-TERM PRESERVATION

- Zenodo (CERN-hosted, EU-funded repository)
- DTU Data (institutional repository)

- EnerMaps and OpenAIRE for enhanced visibility
- Minimum 10-year preservation guaranteed

## SECURITY STANDARDS

### General Data Protection Regulation (GDPR), cybersecurity, and data ethics

Utilising 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. 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 organisations 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.

### 2.1.8. TIMELINE OF DATA COLLECTION AND PRESERVATION

During the SEEDS project execution (January 2024 - December 2027), most research data is subject to an embargo period and is not publicly accessible. Some data will be fully opened in accordance with FAIR principles after the project ends.

#### TIMELINE FOR DATA OPENNESS

- **During project (2024-2027):** Data protected, managed according to FAIR principles
- **When publications are accepted:** Underlying data made open access with a DOI
- **Project end (December 2027):** Metadata for all datasets made openly available (CC0)
- **12 months post-project (December 2028):** Pilot site and operational data made fully open access, unless restricted by the pilot sites

## 2.2. FAIR PRINCIPLES IMPLEMENTATION

FAIR data management is a legal obligation under the Horizon Europe Grant Agreement. Publicly funded research data is considered “a public good” and must therefore be fully opened according to the FAIR Principles after the project ends. Making the data Findable, Accessible, Interoperable and Reusable ensures that the output can be verified, built upon and translated into real-world solutions.

### 2.2.1. MAKING DATA FINDABLE

All SEEDS datasets will be assigned persistent Digital Object Identifiers (DOIs) and described with rich metadata to enable discovery. During the project, metadata is created and maintained. After embargo periods, data and metadata will be published in trusted repositories.

### PERSISTENT IDENTIFIERS

Datasets and research outputs shared publicly will be published in repositories that assign DOIs. (Zenodo and DTU Data assign DOIs to all published outcomes).

- DOIs assigned to all published datasets after embargo
- Version control implemented to track updates
- Software code published with DOIs for each release

### RICH METADATA

SEEDS provides comprehensive metadata for all datasets following the DataCite metadata schema:

- Creator/Author names with identifiers
- Dataset title and description
- Keywords (building energy, heat pumps, renewable energy)
- License information
- Publication date
- Funding: EU Horizon Europe Grant Agreement 101138211

SEEDS data will be organised under project identifiers: Zenodo community "EU-HEU-101138211" or DTU Data project "SEEDS".

## 2.2.2. MAKING DATA ACCESSIBLE

During the SEEDS project execution (January 2024 - December 2027), some of the research data is subject to an embargo period and is not publicly accessible. Data will be fully opened in accordance with FAIR principles after the project ends. This approach:

- Protects intellectual property while partners develop innovations and seek patents
- Allows partners to publish findings before making underlying data public
- Maintains competitive advantage for consortium partners developing market-ready solutions
- Ensures GDPR compliance for building-related data

This embargo approach is permitted under Horizon Europe Grant Agreement § 17 (*Obligation to protect results*) and § 26 (*Open science*).

### PUBLICATION-LINKED DATA

Research data underlying publications is made openly available when publications are accepted.

### 2.2.3. METADATA ACCESSIBILITY

All metadata will be openly available and licensed under CC0 (public domain) as per Grant Agreement requirements at the end of the project. This applies even during embargo periods - metadata describing protected data is public. Repositories use standard web-based access protocols (HTTP/HTTPS). No authentication is required for open data, while secure authentication is set up for consortium-restricted data during the project.

### 2.2.4. MAKING DATA INTEROPERABLE

SEEDS partners use standardised data formats appropriate to their data types:

- Time-series building data: CSV, HDF5
- Sensor data: JSON, CSV
- Building characteristics: IFC, gbXML
- Geospatial data: GeoJSON, Shapefile
- Software/code: Python, MATLAB, Modelica, open formats

When project-specific vocabularies are necessary, mappings to commonly used ontologies are/will be provided. Data includes qualified references to related datasets, publications, and code repositories. External data sources are clearly cited with DOIs.

### 2.2.5. MAKING DATA REUSABLE

#### COMPREHENSIVE DOCUMENTATION

All SEEDS datasets include README files with:

- Methodology and data collection procedures
- Variable definitions and units of measurement
- Quality control procedures and known limitations
- Processing steps and transformations
- File structure and naming conventions
- Related resources (DOIs of publications, code)
- Funding acknowledgement (EU Horizon Europe No:101138211)

#### DATA LICENSES

Data will be distributed under standardised licenses:

- CC0 (Public Domain): Preferred for datasets to allow maximum reuse
- CC BY 4.0 (Attribution): Where attribution is essential
- Open-source licenses (MIT, Apache 2.0): For software and code that can be shared (i.e. no IP protection)

#### QUALITY ASSURANCE

SEEDS implements quality assurance across all partners:

- Standardised file naming conventions
- Version control for tracking changes
- Quality control checks before publication
- Documentation of data provenance and processing steps

### 2.3. IMPLEMENTATION ACROSS PARTNERS

All SEEDS partners implement this framework with support from the DTU team. Partners document their specific data management practices in Sections 3 and 4 of this DMP.

The framework aligns with the EU open science requirements and provides legitimate intellectual property protection, ensuring SEEDS innovations can be protected during development and committing to full data openness after the project.

## 3. DATA MANAGEMENT BY PILOT

### 3.1. DANISH PILOT

#### 3.1.1. PURPOSE OF THE DATA

The Danish pilot collects building performance data from Tech House to enable optimization of heating, cooling, and energy systems integrated with renewable energy sources.

#### 3.1.2. COLLECTED DATA

The Danish pilot implements a comprehensive data collection system through the Nordomatic Control and Technical Systems (CTS):

- **Zone-level data:** 98 zones across 3 floors monitoring room temperatures and CO2 levels
- **Energy consumption:** 17 electricity meters, 22 water meters, and 10 district heating consumption meters
- **Temporal resolution:** Data collected at 1-5 minute intervals, with majority at 1-minute resolution
- **Historical baseline:** Data from Nordhuset (similar building) collected from April 2025 to establish baseline performance
- **Data format:** Standardised format with timestamp, ID, value, and unit fields
- **Update frequency:** Daily automated uploads from Nordomatic CTS system

Figure 1 shows the end-to-end dataflow in the SEEDS-project focused on the CTS-data from the specific building Tech House, in which data is obtained from the Nordomatic CTS-system and made available on the Center Denmark Portal.

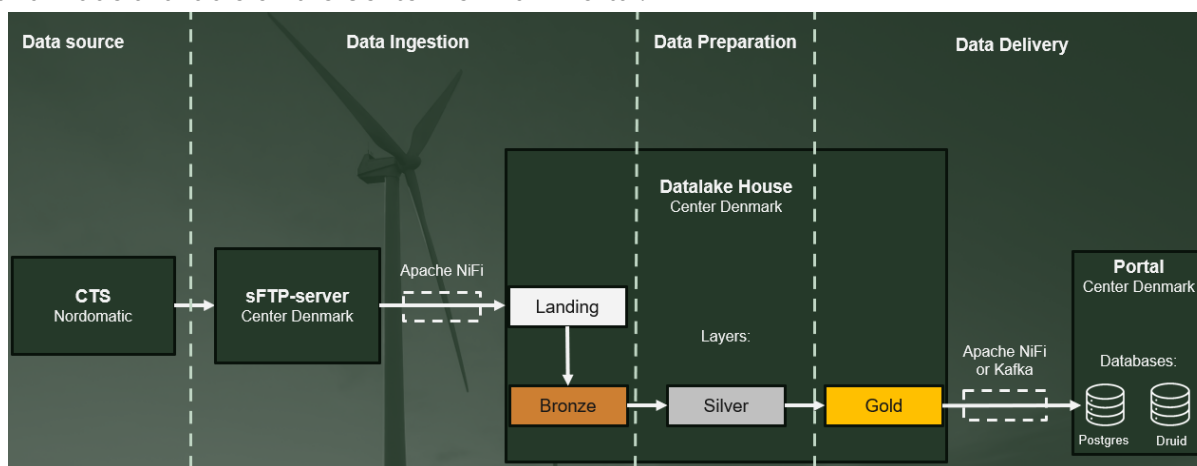


Figure 1, Visualisation of end-to-end data flow

As presented in Figure 1, the collected data undergoes a sophisticated end-to-end processing pipeline:

- **Data ingestion:** CSV files uploaded daily to CDK's SFTP server via Nordomatic CTS
- **Landing zone:** Raw data stored in staging layer for initial processing

- **Bronze layer:** CSV files transformed to delta tables using Apache NiFi for structured data processing
- **Silver layer:** Data quality checks including de-duplication, schema enforcement, and partitioning
- **Gold layer:** Application-ready datasets with enrichment from complementary sources
- **Data enrichment:** Integration with BBR building data, weather data (municipality or 10x10 km grids), and electricity prices
- **Analysis and research:** Researchers access data through CDK Portal with role-based authentication for system optimisation and performance validation. Figure 2 shows the portal provided by CDK.

El	Varme	Varmt Vand	Koldt Vand	
Adresse		Målerstand	Øjeblikseffekt	Beskrivelse
	QM1	0,001 MW-hr	0 L/hr	Skov Advokater
	QM2	39,526 MW-hr	0 L/hr	Lejemål 1
	QM3	35,111 MW-hr	0 L/hr	Lejemål 2
	QM4	8,910 MW-hr	0 L/hr	Lokal Bolig
	QM5	33,465 MW-hr	0 L/hr	Handels Banken
	QM6	53,590 MW-hr	0 L/hr	Seaway7
	QM7	54,036 MW-hr	0 L/hr	Martines Øst
	QM8	38,809 MW-hr	0 L/hr	Martines Vest
	QM9	64,188 MW-hr	0 L/hr	Skov Advokater Øst
	QM10	77,837 MW-hr	0 L/hr	Skov Advokater Vest

Figure 2, Example of district heating metering data from the Nordomatic CTS system

### 3.1.3. HOW DATA IS USED

The data collected by CDK is utilised by DTU for offline model development and real-time control implementation. For data-driven modelling, historical data from the CDK data portal is used to identify, tune, and calibrate models for the various tasks in WPs 2-5. E.g., white-box and grey-box models that characterise building thermal dynamics, HVAC system behaviour, and apply controls to activate energy flexibility. Once calibrated, the models are used within an optimisation and control framework for prediction, energy management, and grid-service planning. In the real-time implementation, high-resolution data is sent to DTU to generate control signals, which are then applied through the building's BMS system.

Alenergy uses the pilot consumption data to support PV system dimensioning and configuration in the Danish pilot within WP2. The measurements are used as input to the SolarPlanner tool to assess feasible PV system sizes and candidate configurations, including storage options. Within WP5, the planned data use supports bidding optimisation using PV

generation time series generated in SolarPlanner, combined with consumption measurements from the CDK portal and ENFOR-provided forecasts.

ENFOR utilises the data collected by Center Danmark to drive operational forecasting systems for solar PV production and electricity consumption, specifically for WP5. Through the Center Danmark API, ENFOR accesses historical power measurements from the TechHouse building. While this metering data is captured at a high frequency (one-minute resolution) and made available once daily for the preceding 24 hours, ENFOR internally resamples these measurements to a 15-minute resolution to match the temporal requirements of the forecasting models. This operational data loop is completed when the generated forecasts are uploaded back onto the Center Danmark platform, making them available for integration with the wider SEEDS systems.

### 3.1.4. HOW DATA IS STORED

The Danish pilot implements a state-of-the-art data lake architecture, ensuring secure, scalable, and accessible data storage:

- **Infrastructure:** Center Denmark Data Lake House with multi-layer architecture (Landing → Bronze → Silver → Gold)
- **Master data:** Stored in PostgreSQL database, managed via Apache NiFi
- **Time-series data:** Stored in Apache Druid database optimised for high-frequency temporal data, managed via Apache Kafka
- **Access control:** The data is made available on the Center Denmark Portal for researchers in the SEEDS project to access the data with role-based access.
- **Security:** SFTP-server as files and stored in the Datalake House of Center Denmark, in the landing zone
- **Backup:** Redundant storage with automated backup procedures
- **Compliance:** Full GDPR compliance with data protection measures
- **Documentation:** Comprehensive metadata and data lineage tracking

### 3.1.5. TIMELINE OF DATA COLLECTION AND PRESERVATION

Raw data from the Danish pilot is collected continuously throughout the project implementation and demonstration. Data access will be restricted to partners within the SEEDS consortium and will only be available during the project execution period (January 2024 – December 2027). The raw datasets will not be made publicly available.

At the end of the project, access to the raw datasets will be terminated, and the raw data will be deleted or archived in accordance with the agreement with the pilot site owner and the data governance procedures of the organizations involved.

Results and derived outputs such as aggregated datasets, KPIs, analysis results, and control-related conclusions will be disclosed according to the SEEDS dissemination procedures and in agreement with the pilot site owner.

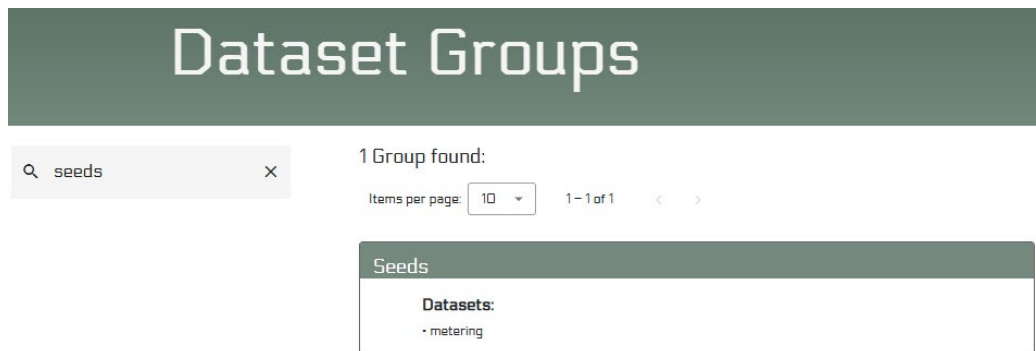
### 3.1.6. FAIR CRITERIA

#### Findable

with Metadata: Upon clicking on the metering dataset, it is possible to read the metadata

#### Accessible

Data users have been granted role-based access to the dataset on the portal and can find the dataset on the Portal – either by groups or on the main page for datasets. See Figure 3.



*Figure 3, Example of dataset management*

#### Interoperable

Interoperability is supported through CDK's infrastructure allowing role-based API access through standardised structures. The pilot data is collected from Nordomatic CTS and delivered in a standard format including timestamp, ID, value and unit fields, enabling systematic integration with analytics tools and external datasets. The CDK data platform supports enrichment and harmonisation of datasets, allowing users to combine building performance data with complementary datasets such as: BBR building registry, weather data, and Electricity price datasets.

#### Reusable

Reusability of the Danish pilot data is ensured through uniform access and quality assurance in the CDK data platform. The datasets are continuously updated at high temporal resolution while preserving historical values, enabling multi-purpose data use and reuse. Concretely, the datasets enable use for both offline model development, calibration, and validation, and real-time operational optimisation and control. The data will be available for continuous reuse throughout the project, e.g., for control performance evaluation, validation, further analysis, and algorithm verification.

## 3.2. SLOVENIAN PILOT

### 3.2.1. PURPOSE OF THE DATA

The primary objective of the Slovenian pilot project is to design, implement, and assess a system that enables the provision of electricity flexibility services. The pilot site comprises five geographically dispersed locations, and the system will exploit devices including heat pumps, water storage tanks, and solar panels to deliver flexibility in response to demand requests from the Distribution System Operator (DSO). These pilot locations are retail points, specifically Petrol gas stations, and we collect a wide range of data on building efficiency. The collected data includes power consumption metrics, various sensor readings, and performance data from local heating and cooling systems, as well as solar power generation and site-specific weather conditions. The data is used for training forecasting models, which are a key component of the system for providing flexible services.

### 3.2.2. COLLECTED DATA

For each of the pilot locations, we collect the following data types:

- Local meteorological data
- Public meteorological data and weather forecasts (provided by the Slovenian Environment Agency)
- Customer number descriptors (including door opening and water consumption)
- Energy consumption of different components and subsystems
- States and settings of different components and subsystems
- Electricity production by a local solar power plant

For example, at Bled location the collected data includes:

- Monitoring of temperature and CO<sup>2</sup> levels in the retail area
- Energy consumption measurement via 6 individual electric meters
- Door status sensor (for monitoring door openings/closings and opening width)
- Ultrasonic occupancy sensor in the store manager's office
- Local meteorological station readings

Data is read at 15-minute intervals and stored in a relational database. The data follows a standardised format, including fields for timestamp, ID, value and unit. For developing initial forecasting models, we also used data from this facility prior to the upgrade, starting from June 2019.

Data is primarily collected via a BMS (Movicon), and partially through individual cloud solutions (Beaver IoT, Petrol IoT). An overview of technologies used for data collection is presented in Table 3.

Table 3, Technologies used for data collection at different system levels

Layer	System	Protocol / Method
Field Level		Sensors, Inverters, HVAC
Integration Level	Movicon (BMS)	OPC UA / MQTT
Cloud Level	Beaver IoT, Petrol IoT	REST API / HTTPS

### 3.2.3. HOW DATA IS USED

The data collected by Petrol in a database is automatically mirrored every hour to a database within JSI premises. The data is used for training forecasting models and assessing the state of each individual location and its capacity to provide on-demand flexibility services. The data is a crucial resource for the development and operation of the flexibility service. Data processing is performed within a secure computing environment.

### 3.2.4. HOW DATA IS STORED

The data is processed and stored by two consortium partners – Petrol and JSI. Petrol uses an on-premises MS SQL database for data storage, and the TimeScale Cloud solution for time series data.

MS SQL database backups are performed daily in a local datacenter; TimeScale database is backed up as part of a cloud solution. Data storage is fully compliant with GDPR data protection measures. JSI also uses an MS SQL database to store the data; the database is located within the JSI secure computing environment.

### 3.2.5. TIMELINE OF DATA COLLECTION AND PRESERVATION

- June 2019 – September 2025: Collection of data from the original (non-renovated) Bled location
- September 2025 – December 2027: Collection of data from renovated locations

At the end of the project, JSI's access to data will be terminated, and JSI will delete or archive the data in accordance with the NDA agreement.

### 3.2.6. FAIR CRITERIA

Except for the public meteorological data, all collected data is regarded as a business secret by Petrol and is available to other project partners only upon signing a separate nondisclosure agreement (NDA). For the purposes of the project, it is planned that the only partner needing access to data is JSI, and a suitable NDA has been signed by Petrol and JSI at the start of the project.

### Findable

The data is provided with an accurate title and rich metadata. Standardised tags, timestamps, and keywords are used.

### Accessible

Data is accessed via a platform over HTTPS protocol where the user authenticates via Keycloak and we have a defined user role that determines the level of access and the set of data to which the user is entitled. This prevents unauthorized access, while ensuring availability to authorized users. Access is also possible via API, where access is verified in two levels.

### Interoperable

Interoperability is ensured by using the JSON format, which is easily read by various programs. Standardized notations are used for units, dates, etc. The connections between data are described in a way that is also understood by computer algorithms.

### Reusable

The data is documented, versioned, and time stamped. For more efficient indexing, the data also includes appropriate metadata. The data is also stored in RAW format.

## 3.3. BELGIUM PILOTS

### 3.3.1. PURPOSE OF THE DATA

The Belgian pilot and replica site collect building performance data from sensors and components of their local district with collective heating (and cooling), representing two separate neighbourhoods in the city of Bruges, in two distinct geographical locations. After hybridisation and installation of a switchable hydraulic configuration, the goal is to implement and assess an anticipative automated control, enabled by Model Predictive Control (MPC), that optimises energy efficiency (or cost) and thermal comfort. The first pilot (De Schipjes) is a Living Lab that comprises 12 single-adult households (data collection ongoing). The second pilot (Stijn Streuvel) is a replicator that comprises 15 single-adult households (currently under renovation). The data is collected using a building management system, provided by Priva for de Schipjes and by Schneider for Stijn Streuvel. Some additional LoraWan sensors were installed. The data and BACnet network are subsequently used by MPC to calculate and send optimised control signals to the heating/cooling system, to optimise the use of the fully-renewables-based hybrid collective energy systems on site. All energy systems use renewable energy sources: air-water heat pumps, water-water heat pumps connected to a (zoned) borefield, solar thermal collectors, PV panels, and PVT panels.

Moreover, simulations and optimisations are carried out for both pilots using detailed physics-based models in Modelica to test (different versions of) MPC, integrated optimal control and sizing (IOCS) methods and a novel fault detection method. The resulting data is used to assess these methods.

### 3.3.2. COLLECTED DATA

The following measurement data is collected in the De Schipjes Pilot:

- Electrical meter energy:
  - Total electricity of the energy cabin (central heating system)
  - Total electricity per house
  - Booster heat pump electricity per house
- Control signal of all actuators (pumps, valves) that are controlled via the Priva BMS
- Thermal energy of the heating network components:
  - The main district heating network:
    - Temperature (inlet/outlet)
    - Flow rate
    - Thermal energy
  - Each house heat exchanger:
    - Temperature (inlet/outlet)
    - Flow rate
    - Thermal energy
  - Ground source heat pump condensor
    - Flow rate
    - Thermal energy
  - Ground source heat pump evaporator
    - Temperature (inlet/outlet)
    - Flow rate
    - Thermal energy
  - Air-water heat pumps condensers
    - Temperature (inlet/ outlet)
    - Flow rate
    - Thermal energy
  - Viessmann solar thermal collectors
    - Thermal energy
    - Flow rate
  - Geothermal borefield, original 8 boreholes
    - Thermal energy
  - Geothermal borefield, 2 new boreholes
    - Temperature (inlet/outlet)
    - Flow rate
    - Thermal energy
  - Free cooling heat exchanger secondary side:
    - Temperature (inlet/outlet)
    - Flow rate
    - Thermal energy
- Electrical energy consumption of heat pumps: to be installed
- Fiber optic temperature measurements of the two new boreholes (distributed temperature sensor system (DTS))- continuously
- Hot and cold water use for each district house
- Temperature sensors at each house thermostat and/or in additional rooms. (LoraWan)

- Solar irradiation sensors (LoraWan)
- dB sensors to measure noise levels of AW HPs: to be installed
- Some major alarms
- Other temperature sensors in the system
  - Buffertanks (inlet/ outlet)
- All relevant setpoint temperatures
  - Heat pumps
  - Heat exchangers
- Buffer tanks

All of the data entries listed above will be measured in the replica site Stijn Streuvels. The following additional measurement data will be collected in the Stijn Streuvel Pilot:

- Each borefield zone has one borehole with a continuous fiber optic temperature measurement (distributed temperature sensor system (DTS))

This will be completed when all sensors have been installed (renovation is ongoing).

After all simulations and optimisation runs, all relevant physical variables computed by the pilot models are systematically collected. The models themselves are managed via GitLab, ensuring version control and traceability.

Datasheets of installed components are collected on the SWECO Sharepoint.

### 3.3.3. HOW DATA IS USED

MPC uses relevant data to compute real-time optimal control set points for the energy system. Moreover, they are inputs for MPC model calibration and (automated) fault detection.

Minimal processing ensures data integrity by storing read errors as 'none' values rather than omitting them. Data is transmitted securely over HTTPS to a central cloud storage server (Builtwins). Data can be accessed via a token-secured HTTPS API (providing JSON-formatted data for automation) or through Grafana, which enables users to visualise, analyse, and export data in CSV format. KU Leuven uses Grafana to export measurement data for archiving in ManGO. These data are used for model validation, to gain deeper insight into geothermal regeneration processes (observable via optical-fibre measurements), to support the development and tuning of state observers for MPC, to assess system performance and to analyse sensor behaviour, data quality, and the impact of real operational disturbances on system and control performance. Processing methods vary based on application needs: real-time monitoring may interpolate missing data, calibration models may ignore it, and Model Predictive Control (MPC) can replace it with the last known value. This flexibility ensures reliable data-driven decision-making for analytics, system calibration, and optimisation.

Data generated during simulations and optimisation runs are used to evaluate the performance of the developed MPC (multiple versions) and IOCS methods. MPC model outputs are confronted with measured data in the (automated) fault detection.

### 3.3.4. HOW DATA IS STORED

Data is stored at equidistant intervals: every minute for the first 100 days, then every 15 minutes for up to six years. Older data is aggregated using methods such as averaging to optimise storage without compromising analytical value. More information is provided on the graphite time series database documentation (<https://graphiteapp.org/>).

Telemetry data from various building systems—such as room temperature sensors, energy meters, HVAC systems using the existing Priva control platform, and IoT devices—is collected and processed by a local on-site device. The collection occurs via BACnet protocols, either through polling (reading values at intervals) or subscription-based updates (triggered when a threshold is exceeded). For energy meters and some other devices that use Modbus or M-bus, the data is accessed through BACnet where the main Priva controller serves as a translation unit. The data is uploaded to the Builtwins cloud server Graphite database via the aforementioned API services. The Builtwins server acts as the central data hub.

A subset of the available sensor data is also downloaded daily and stored using ManGO (<https://rdm-docs.icts.kuleuven.be/mango/index.html>). This is the KU Leuven Research Data Management platform for secure, structured storage, version control, and metadata for research data. The pipeline ingests data from the Builtwins API and writes a single Apache Parquet file per 24-hour period. Each Parquet file carries designated flow-level and dataset-level metadata that capture ingestion provenance (e.g., run identifiers, timestamps, source system, project, and site), enabling traceability of the archived data. Files are first deposited by the ingress machine account into /TheSySi/ingress and a ManGO flow moves them to the project structure under /TheSySi/projects/20242028\_HorEu\_SEEDS/raw/<site>/<year>/ (processed outputs live under the matching processed/ path), keeping raw and processed data separated and year-partitioned.

The status of De Schipjes shows 2,162 backed-up datapoints, comprising 1,310 measurement datapoints (including 818 fibre optic data points) and 852 MPC outputs. For each BACnet metric identifier, all associated attributes are documented in a designated metrics file. The attributes (e.g., human-readable description, legacy alias, device identifier, BACnet point type and index, point name, building/site code, and source type) are documented in a tag schema ensuring consistent interpretation, unambiguous mapping to control points, and reuse across analysis workflows and sites. Exceptional events affecting data quality (such as sensor downtime, firmware updates or control-equipment interventions) are captured through explicit metadata flags on the affected Parquet files, including timing and short descriptions, to support correct interpretation during analysis. Access to the ManGO project is managed through role-based groups, and the ingestion workflows and job configurations are maintained in GitLab, allowing additional pilots to be added with minimal configuration effort while preserving reproducibility and interpretability across sites.

Once measurement data are available from the Stijn Streuvel pilot, they will also be uploaded to the Builtwins cloud server's Graphite database and added to the same KU Leuven ManGO platform via a pipeline using the Builtwins API.

Data produced through simulations and optimisation procedures are saved in .mat, .csv, or .txt formats and are stored on the KU Leuven SharePoint using an equidistant 1-hour time grid.

Datasheets for the installed components are stored on Sweco's SharePoint, shared among all Belgian pilot partners.

### 3.3.5. TIMELINE OF DATA COLLECTION AND PRESERVATION

- April 2024 - current: temperature logging of the households in De Schipjes
- April 2025 - current (January 2026): temperature logging of the households in De Schipjes + data points listed in Section 3.3.2
- January 2024 - current (January 2026): storage of simulation and optimisation data.

At the end of the project, the (raw) measurement data will be archived in accordance with the agreement with the pilot site owner, and the simulation and optimisation data will be archived in accordance with the data governance procedures of the organisations involved.

Results and derived outputs, such as aggregated datasets, KPIs, analysis results, and control-related conclusions, will be disclosed in accordance with the SEEDS dissemination procedures and in agreement with the pilot site owner.

### 3.3.6. FAIR CRITERIA

#### Findable

All the measurement data collected at De Schipjes are uniquely identified through standardised BACnet annotations. The model developed for MPC uses the BACnet points available throughout the Building Management System to identify the necessary control points and access points. The BACnet data points are identified by name and description. When the name and description are not sufficiently clear, we use the Building Management System (BMS), which visualises those points to map the distinct control points. For Stijn Streuvel the system still needs to be set up.

In ManGO, archived daily Parquet files include searchable file-level metadata (e.g., project, building, timeseries start/end) and each BACnet metric is stored by its unique full identifier while the complete per-metric attributes (description, alias, device, point\_type, point\_index, point\_name, building, source\_type) are available in the designated metrics file following the tag schema.

Simulation and optimisation data stored on the KU Leuven SharePoint follow a standardised folder structure and consistent file-naming conventions, enabling users to easily locate the required files.

#### Accessible

Access to the measurement datasets (both for MPC and for Priva in case of De Schipjes) is granted through authenticated APIs and secured endpoints, with standardized user roles and permissions throughout the datasets, ensuring availability to authorized users and preventing

unauthorised access. In ManGO, role-based permissions, along with recorded file ownership and ingest provenance (e.g., job id, pipeline id, run id, and timestamps), enable authenticated retrieval and an auditable trail for archived files. The simulation and optimisation data are stored on the KU Leuven SharePoint platform, where access is managed through institutional authentication and user permissions, ensuring that authorized users can reliably retrieve the data.

### Interoperable

All measurement data is stored using their full BACnet names, so it can be uniquely identified. Additionally, a large subset of the data is visualised on the Grafana dashboard, grouping the data in logical panels. Data from the Grafana database can be extracted as CSV files or plots. In ManGO, data are stored as Apache Parquet with embedded dataset and flow metadata and a tag-schema reference. The simulation and optimisation data are stored in widely supported, non-proprietary or commonly used file formats (.csv, .txt) or established scientific formats (.mat), allowing seamless use across different software tools and analysis environments.

### Reusable

The identification of the available data stored in the central database aligns with the first mapping of all BACnet points when developing the initial model for MPC. The identifier of each BACNET point refers to a specific control point; consistencies in annotations are crucial to ensure the functioning of MPC for a given site and a given technical installation, in this case De Schipjes. In ManGO, each Parquet file carries provenance (mgs.flow.\*), dataset descriptors (mgs.dataset.\*), and any explicit event/data-quality flags (e.g., sensor downtime or firmware/control-equipment updates), allowing downstream users to interpret and reuse the archived data.

All simulation and optimisation datasets are generated on an equidistant 1-hour time grid and are accompanied by consistent variable definitions, ensuring that the data can be readily understood, reproduced, and reused for future analyses and method comparisons.

## 3.4. GREEK PILOT

### 3.4.1. PURPOSE OF THE DATA

The EL pilot collects building performance, energy, and environmental data from the C1 student dormitory building at Democritus University of Thrace with the explicit purpose of enabling closed-loop, data-driven control of the building's energy assets. The data are used to support the operation of advanced control strategies that calculate and dispatch optimized control signals to on-site energy systems in real time. Data collected from room-level sensors, building-level meters, renewable energy systems, and environmental monitoring are processed by SEEDS analytics and AI-based control services, such as the smart modulation of MSHP operation and the adaptive control of the smart reflector system to maximize solar harvesting (T4.1), or the deployment of black-box MPC (T4.3), to optimize energy efficiency, thermal comfort, and renewable energy utilization.

### 3.4.2. COLLECTED DATA

The Greek Pilot collects time-series sensor data from the C1 building and its associated energy systems. The datasets include:

- Room-level sensor data from selected (20 out of 68) rooms, including:
  - Indoor temperature (°C)
  - Relative Humidity (%)
  - CO<sub>2</sub> concentration (ppm)
  - Electrical energy consumption
  - Thermal Energy
  - Domestic Hot Water (DHW) usage
- Environmental (campus weather station) data, including:
  - Outdoor temperature (°C)
  - Relative Humidity (%)
  - Wind Speed (m/s)
  - Solar radiation (W/m<sup>2</sup>)
- Building-level aggregated data, including:
  - Total electrical power and energy
  - Photovoltaic (PV) power and energy
  - Aggregated DHW consumption
- Energy-assets specific data, including:
  - Multi-source heat pump operational readings (e.g., power, source-mode, etc.)
  - Solar panel sensors readings
  - EV charger power and energy consumption

All these datasets are represented in an NGSI-compliant entity and attribute way, enriched with timestamps generated at ingestion time.

Control signals to the actuation devices are also being collected and stored. The signals include:

- Fan coil signals inside the 20 dwellings (e.g., setpoint, mode, fan speed)
- MSHP signals (e.g., operational mode adjustment)
- PV reflection tracker signals (e.g., reflector distance)

### 3.4.3. HOW DATA IS USED

The collected data from the CERTH IoT platform within the SEEDS are used to:

- Feed near real-time monitoring dashboard and visual analytics tools (T3.4.1). See Figure 4 and Figure 5.

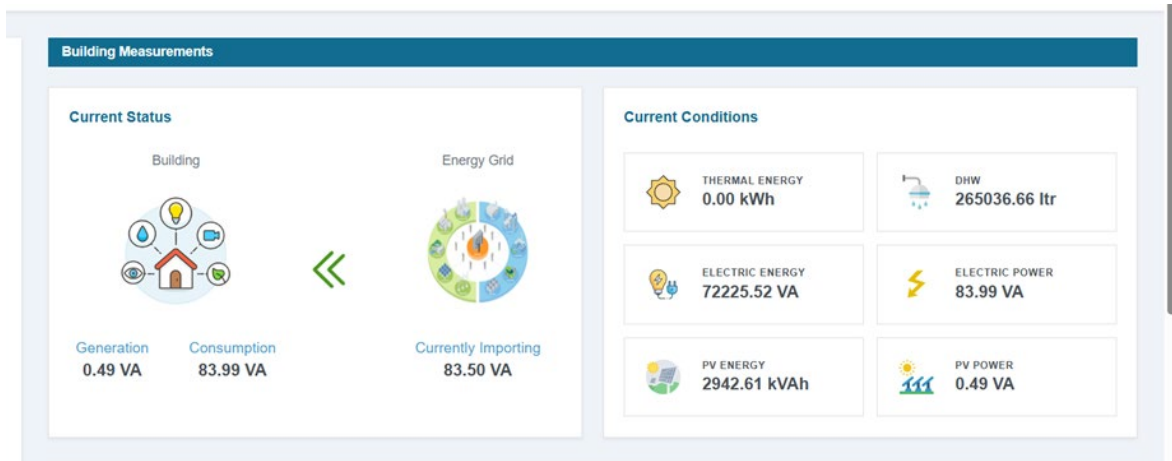


Figure 4, Visual Analytics Engine for facility managers in EL pilot

- Provide notifications and alerts to facility managers

**Notification Details** ×

**Building high electrical demand** resolved

Friday, January 23, 2026 at 4:25:59 PM GMT+02:00 📍 Building:C1

**Summary**

High electrical demand detected at building level. Current demand is 72.63 kW, exceeding the configured threshold.

**Alert Details**

<b>Rule:</b> BuildingHighDemand	<b>Threshold:</b> 60	<b>Trigger Value:</b> 72.63
<b>Severity:</b> <b>WARNING</b>		

× Close

Figure 5, Notifications and Alerts for facility managers in EL Pilot

- Train and validate AI models for:
  - Smart modulation control of multi-source heat pumps (T4.1)
  - Smart tracking and energy yield optimization of PV energy systems (T4.1)
  - Predictive maintenance of electrification assets (T4.4)
  - Support thermal flexibility forecasting, proactive demand reshaping, and grid-support services (WP5)

- Evaluate the performance and impact of SEEDS innovations under real operational conditions

Data processing and analytics are performed exclusively within the SEEDS consortium and by authorized project partners. Data are being ingested from the pilot devices and energy assets to a deployed DUTH master server, and are then forwarded to the CERTH IoT platform every five minutes using a standardized HTTP-based RESTful API. This platform is being used both as the main data collection entry point for all the consumers of EL Pilot data and as the main remote controller point of the pilot's energy assets.

#### 3.4.4. HOW DATA IS STORED

Data is ingested and managed using a FIWARE-based architecture, composed by some of the following relevant software components:

- FIWARE IoT Agent (JSON): used for device and sensor integration
- Orion Context Broker: used for semantics, alert notifications, context data management and current system state
- QuantumLeap: used for time-series data persistence
- CrateDB for scalable storage of historical time-series data
- Mysql database: used for storing user account data (roles, permissions etc.)

Access to all the datasets is protected through dedicated PEP proxies using Keyrock as the Identity Manager component of the system, implementing role-based access control and authentication mechanisms. Historical data is stored in servers operated by CERTH, following all the relevant institutional compliance policies.

The developed backup data plan foresees the creation of periodic snapshots of all databases operating within the IoT platform. These snapshots are timestamped and securely stored on a bi-monthly basis, ensuring data integrity, traceability, and recoverability in case of system failures, data corruption, or other unforeseen incidents.

#### 3.4.5. TIMELINE OF DATA COLLECTION AND PRESERVATION

Data collection in the Greek pilot starts during the deployment and commissioning phase of the SEEDS project and is performed at a fixed interval of every five minutes. The collection process remains active throughout the full operational lifetime of the Greek pilot, ensuring continuous availability of time-series data for monitoring, analysis, and validation activities. The collected data are retained for the entire duration of the SEEDS project, in order to support reporting and evaluation processes. Upon the completion of the defined retention period, the data will either be securely archived or permanently deleted in accordance with the preservation policies defined by the data owner (DUTH).

#### 3.4.6. FAIR CRITERIA

The data collected in the Greek pilot adheres to the FAIR principles as follows:

### Findable

All the data collected in the Greek Pilot are uniquely identified through standardized NGSII entity IDs and metadata, allowing efficient discovery within the SEEDS data pool. See Figure 6

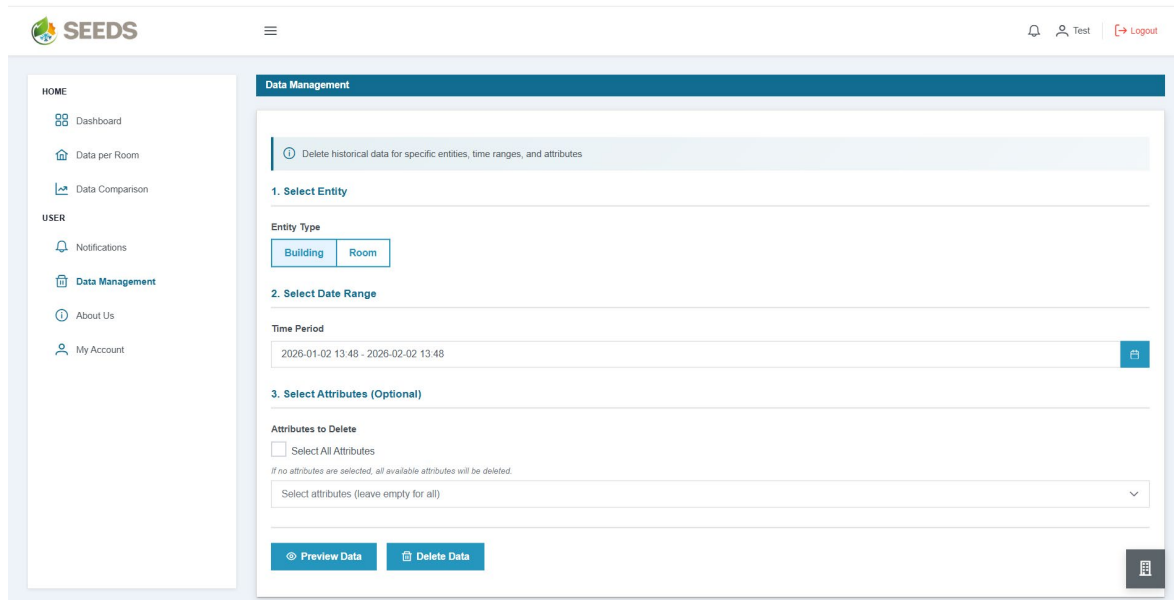


Figure 6, Data Management feature for facility managers in EI Pilot

### Accessible

Access to the datasets is granted through authenticated APIs and secured endpoints, with standardised user roles and permissions throughout the datasets, ensuring availability to authorised users and preventing unauthorised access.

### Interoperable

All the data is structured using NGSII standards FIWARE-based data models, enabling seamless interoperability with all the SEEDS components and external tools

### Reusable

Data from the Greek Pilot are well-documented, timestamped and stored with consistent units and semantics, allowing reuse for analytics, tool validation and research activities within the limits of the SEEDS project.

## 4. DATA MANAGEMENT BY WORK PACKAGE

---

### 4.1. WP1 DATA MANAGEMENT

#### 4.1.1. PURPOSE OF THE DATA

All data related to WP1 “project management and coordination” is collected to support the project process. Some data is collected and used to assess the process and progression of the project, while other data focuses on communication. Additionally, the data helps project partners track progress, upcoming tasks, and deadlines. Overall, the collected data has the purpose of:

- Project management
- Innovation management
- Quality assurance
- Risk management
- Legal, knowledge and data management

#### 4.1.2. COLLECTED DATA

The types of data collected for data management are diverse. For each management task, the relevant data collection is listed below:

##### Project management

The data related to project management includes:

- Contact information, e.g., names, partners, email addresses and roles in SEEDS
- Status of milestones and deliverables
- Meeting minutes
- Communication with the PO, and collection of required/missing information

##### Innovation management

The innovation management is considered in the delivery D1.3. The content is sensitive and will not be described in detail here. For consortium members, the file is available [here](#).

##### Quality assurance

The quality is addressed in the Project management plan (D1.1), the innovation management plan (D1.3) and the risk management plan (D1.4). These will, however, not be available externally, due to sensitive content.

##### Risk management

The data collection related to risk management primarily involves continuous collection of risk status, mitigation plans and mitigation actions from all partners.

##### Legal, knowledge and data management

The data collected consists of non-personal, non-sensitive information related to technologies, business concepts, pilot installations, and market pathways. No personal data as defined under GDPR (Regulation (EU) 2016/679) is collected, processed, or stored beyond the duration of the project. The legal considerations mainly concern IPR, confidentiality and compliance.

- Intellectual Property Rights (IPR):
  - o Data includes IPR strategies of project results. Ownership, access rights, and exploitation conditions are governed by the Consortium Agreement and relevant Grant Agreement provisions.
- Confidentiality:
  - o Certain data related to business strategies, cost structures, or market positioning are treated as confidential and shared only within the consortium under agreed confidentiality obligations.
- Compliance:
  - o Data collection and use comply with Horizon Europe rules on dissemination, exploitation, and protection of results. No regulatory approvals or ethical authorisations are required for the data collected.

#### 4.1.3. HOW DATA IS USED

Generally, data is used to track the project's progress and to inform partners what to focus on at any given time. They are also used to knowing when to intervene in a task, deliverable or milestone to facilitate the best conditions for SEEDS as a whole.

#### 4.1.4. HOW DATA IS STORED

Generally, the data is stored in the SEEDS Teams folder, as it serves not only those involved in WP1, but the whole consortium. A few things are kept in a Teams folder for the DTU management team, to keep track of more sensitive information and ensure that nothing is deleted or moved by accident. Most data are stored and dynamically changed as needed in developed Excel-sheets.

#### 4.1.5. TIMELINE OF DATA COLLECTION AND PRESERVATION

The collected data in WP1 supports the process of SEEDS; however, not all of it is considered an actual outcome of the project. E.g., dynamic lists of contact information, and meeting minutes will not be publicly available.

#### 4.1.6. FAIR CRITERIA

The data collected for WP1 is fundamental to the project process, but it should not be considered an outcome of the project. Thus, the compliance with the FAIR criteria is limited. Internally in SEEDS, the data are, however, managed in a structured way, which would make it possible for others to take over the management.

##### Findable

All management data is kept in organised Teams folders

##### Accessible

The data are available for those with granted access to the Teams folders

##### Interoperable

The data are kept in ordinary file formats, e.g., Excel and Word

## Reusable

Templates and Excel files can be used for other projects; however, they are tailored to support the aims, goals, and progress of SEEDS.

## 4.2. WP6 DATA MANAGEMENT

### 4.2.1. PURPOSE OF THE DATA

The purpose of the data collection is to support WP6 “Replication strategies, exploitation and business models” activities related to exploitation, business modelling, market readiness, and replication strategies.

Data are collected to:

- Characterise and assess the exploitable results generated within the project.
- Support the definition of viable business models and business cases.
- Identify pathways, costs, and milestones required to reach TRL9 and market uptake.
- Enable the design of internal and external replication strategies based on pilot solutions.
- Ensure consistency, traceability, and validation of exploitation-related decisions through partner engagement.

The data ultimately contribute to project deliverables related to exploitation strategy, business models, TRL9 roadmap, and replication frameworks, supporting both internal decision-making and dissemination outcomes.

### 4.2.2. COLLECTED DATA

The project collects qualitative and semi-structured data through dedicated tools and questionnaires, including:

#### Task 6.1 – Exploitation Strategy

Data collected via the Characterisation Table, a structured descriptive tool used for each Exploitable Result (ER), including:

- Technical description and innovation content
- Competitive advantage and differentiation
- Technology Readiness Level (TRL)
- Target market analysis (customers, partners, competitors, time to market)
- Intellectual Property Rights (IPR) management considerations
- Revenue streams and cost structures

These data are periodically reviewed, updated, and validated through one-to-one interactions with project partners.

## Task 6.2 – Business Models and Business Cases

Inputs from the Characterization Tables are complemented by a Business Model Canvas Questionnaire, collecting data on:

- Value proposition
- Key partners, activities, and resources
- Customer segments and relationships
- Cost structure and revenue logic

Data are iteratively refined throughout the project lifecycle.

## Task 6.3 – TRL9 Roadmap

Data are collected through a TRL9 Questionnaire, focusing on:

- Technical, organisational, and regulatory steps required to reach market readiness
- Estimated timelines and milestones
- Cost components associated with scale-up and market entry

## Task 6.4 – Internal Replication Strategy

Data collection focuses on:

- Technical and operational characteristics of pilot plants developed within the project
- Benefits and performance features of pilot solutions
- Requirements and needs of internal replicators (entities within the consortium)
- Assessment of the match between pilot solutions and replication needs

## Task 6.5 – External Replication Strategy

A similar approach (vs T6.4) is adopted, with data collected on:

- Pilot solution characteristics
- Needs and constraints of external replicators outside the consortium
- Feasibility and transferability of replication in different contexts

### 4.2.3. HOW DATA IS USED

The collected data are used to:

- Develop and update exploitation strategies and business cases.
- Inform strategic discussions within the consortium.
- Support evidence-based decision-making on market readiness and replication potential.
- Produce project deliverables, reports, and summaries for dissemination.
- Enable consistency across Tasks 6.1–6.5 by using harmonised data inputs.

Data are analysed, synthesised, and aggregated, not used for automated decision-making or personal profiling.

#### 4.2.4. HOW DATA IS STORED

All data are stored in the shared project repository accessible to the consortium partners. Data storage characteristics:

- Digital format (documents, spreadsheets, questionnaires).
- Access restricted to authorised project participants.
- No storage of sensitive personal data.
- Version control applied where relevant during the project lifetime.

No external commercial databases or third-party cloud services outside the consortium framework are used.

#### 4.2.5. TIMELINE OF DATA COLLECTION AND PRESERVATION

- Data collection period: Month 1 to Month 48 (entire project duration).
- Update frequency: Iterative updates during the project lifetime.
- Post-project phase: Data will no longer be updated after project completion.
- Preservation: Data will be retained in the project repository according to consortium and funding programme requirements.

#### 4.2.6. FAIR CRITERIA

The data management approach follows the FAIR principles as follows:

##### Findable

Data are organised in a structured project repository, with clear file naming conventions and task-based folders.

##### Accessible

Data are accessible to authorised consortium partners through the shared project environment.

##### Interoperable

Data are collected using standardised templates and questionnaires, enabling comparison and aggregation across tasks and partners.

##### Reusable

Data are documented, versioned, and structured to allow reuse for reporting, replication strategies, and future research or innovation activities, subject to consortium agreements.

## 4.3. WP7 DATA MANAGEMENT

### 4.3.1. PURPOSE OF THE DATA

The data related to WP7 “External replication framework” is, unlike the data collected in the pilots, non-technical data. The objective of this work package is to determine barriers and provide solutions to accelerate the external replication of SEEDS’ solutions outside of the pilot projects. To this end, interviews were conducted with decision makers. T7.4 focuses on demonstration of added value of SEEDS’ solutions, by means of KPIs collected in the pilots.

### 4.3.2. COLLECTED DATA

Three main data types are distinguished:

- Contact information of decision makers
  - Contact information, e.g., names, partners, email addresses and roles of interview candidates (table)
- Interview outputs
  - Transcribed recordings of the interviews (unstructured data)
  - Analysis of interviews (unstructured data)
- Pilot KPIs
  - Retrieved from WP3 and data collection processes in pilots. The reader is referred to the previous sections.

### 4.3.3. HOW DATA IS USED

The decision maker contact information is used for scheduling the interviews and follow-up afterwards. The interview outputs are used to come up with a problem statement definition, followed by a toolbox with solutions and accelerators. The pilot KPIs are integrated in the visual storytelling to convince decision makers of added value.

### 4.3.4. HOW DATA IS STORED

The decision maker contact information is stored in an Excel file on the DTU SharePoint, where it is stored for the duration of the SEEDS project. The interview recordings are stored for a one year period in OneDrive (Teams recording). The transcriptions are stored for the duration of the SEEDS project on SharePoint. The analysis of the interview outputs is conducted in Word and Excel files, all on SharePoint. These are stored for the duration of the SEEDS project, and published in a public deliverable *D7.1 Problem Statement Definition Report*. For the pilot KPIs, see other sections.

### 4.3.5. TIMELINE OF DATA COLLECTION AND PRESERVATION

The decision maker contact information is collected over the course of 2025 and stored until end of project. The Interview outputs are collected over the course of 2025 and stored until end of project (exception: recordings, one year preserved). For the pilot KPIs, see previous sections.

#### 4.3.6. FAIR CRITERIA

The decision maker contact information is not meant to be shared externally. The interviews analysis and withheld KPIs will be shared publicly in the report and SEEDS webpage, respectively. This data is unstructured and non-technical. Hence, interoperability and reusability criteria are not applicable.

### 4.4. WP8 DATA MANAGEMENT

#### 4.4.1. PURPOSE OF THE DATA

Data related to WP8 are collected to support and evaluate the SEEDS project's communication, dissemination, and stakeholder outreach processes. A primary objective is to track the impact and market penetration of project outcomes through targeted communication activities to ensure rapid uptake of results. Furthermore, this data management enables effective outreach to the public, investors, and researchers, ensuring optimal knowledge transfer and collaboration across sectors. Finally, it enables monitoring the performance of communication tools and dissemination channels through specific Key Performance Indicators.

#### 4.4.2. COLLECTED DATA

The data types collected within WP8 are diverse and focus on project visibility and stakeholder engagement. This includes digital audience metrics such as statistics on website performance and social media engagement on LinkedIn and X. WP8 also maintains a Dissemination Network database, which is a GDPR-compliant contact list including names, organisations, and email addresses of potential early adopters. Other data types include communication collaterals like branding materials, brochures, and project videos, as well as scientific records such as metadata for peer-reviewed journal publications and international conference presentations.

#### 4.4.3. HOW DATA IS USED

Data is used to assess the project's reach and to inform future dissemination strategies by updating the Plan for Dissemination, Exploitation, and Communication (PDER) annually. Contact information is utilised to invite stakeholders to webinars, local engagement events, and the project's workshops. These performance indicators also help assign responsibilities and track the effectiveness of various dissemination channels throughout the project lifecycle. Lastly, publication data is used for continuous technical reporting.

#### 4.4.4. HOW DATA IS STORED

Internal working documents, including operational lists, contact tables, and draft reports, are stored on the project's secure SharePoint and Teams platforms administered by DTU. To protect sensitive information, a dedicated Teams folder is used specifically for WP8 leaders. Public research data, such as final peer-reviewed scientific publications and their underlying

datasets, are deposited in trusted repositories like Zenodo, or institutional repositories such as DTU Data.

#### 4.4.5. TIMELINE OF DATA COLLECTION AND PRESERVATION

Data collection is continuous throughout the 48-month project duration, starting from January 2024. The project website and its associated performance data will stay active for two years after project closure to facilitate continued exploitation of the results. Scientific publications and associated research outputs will be preserved in institutional repositories for a minimum of ten years. Personal contact data in the Dissemination Network will be handled according to GDPR retention policies and deleted or anonymised after project completion unless explicit consent is provided for further contact.

#### 4.4.6. FAIR CRITERIA

WP8 data management adheres to the FAIR principles to ensure that all research outputs are findable, accessible, interoperable, and reusable. All public deliverables and scientific publications are assigned persistent Digital Object Identifiers (DOIs) to ensure they are findable. These publications are made available via immediate open access in trusted repositories, fulfilling the accessibility requirement. Interoperability is ensured by using standardised formats such as PDF for reports.

## 5. CONCLUSION

---

In conclusion, the SEEDS project is committed to exemplary data management practices that prioritise 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 the design, construction and operational phases, as well as from building and urban levels.

The project emphasises the importance of privacy and informed consent, implementing rigorous data handling protocols to protect personally identifiable information. Data will be anonymised where necessary, and access to sensitive information will be strictly controlled. As shown in the data management plan details by pilots in section 3 and by work packages in section 4, the SEEDS consortium will utilise 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 for Horizon Europe projects, 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



[project-seeds.eu](https://project-seeds.eu)



[SEEDS EU project](#)



[@EU\\_SEEDS](#)



Co-funded by  
the European Union

This project is co-funded by the European Union's Horizon Europe  
Innovation actions programme under the Grant Agreement n°101138211