



TWINVEST

Digital Twin for Forecasting of Power
Production to Wind Energy Demand

Project title	Digital Twin for Forecasting of Power Production to Wind Energy Demand		
Project acronym	TWINVEST		
Project number	101146936		
Project start date	01/06/2024	Duration	42 months

D9.3 – DATA MANAGEMENT PLAN

Due date	30/11/2024	Delivery date	30/11/2024
Work package	WP9		
Responsible Author(s)	Foad Gandoman (RSTER),		
Contributor(s)	Name (Beneficiary Name), Name (Beneficiary Name)		
Reviewer(s)	Name (Beneficiary Name), Name (Beneficiary Name)		
Version	V01		
Dissemination level	Public		

VERSION AND AMENDMENTS HISTORY

Version	Date (MM/DD/YYYY)	Created /Amended by	Changes
V00	15/11/2024	16/11/2024	16/11/2024
VO1	20/11/2024	20/11/2024	25/11/2024







TABLE OF CONTENTS

TABLE OF CONTENTS.....	3
LIST OF ABBREVIATIONS	4
1 EXECUTIVE SUMMARY	5
2 OBJECTIVES	5
3 INTRODUCTION.....	5
4 DATA MANAGEMENT PLAN	5
4.1 DATA MANAGEMENT PLAN ON TWINVEST	6
4.1.1 Data Management Methodology.....	6
4.1.1.1 Data type and how to reference.....	6
4.1.1.2 Metadata in Data Management Register Tool	7
4.1.2 Data collection	8
4.2 FAIR DATA	8
4.2.1 Making Data Findable, including provisions for Metadata	8
4.2.2 Making Data Openly Accessible	9
4.2.3 Making Data Interoperable	9
4.2.4 Increase Data Re-Use	10
4.3 ALLOCATION OF RESOURCES	10
4.4 DATA SECURITY	10
4.5 ETHICAL AND LEGAL ASPECTS	11
5 RESULTS	11
6 CONCLUSION	11
7 REFERENCES	12

LIST OF TABLES

TABLE 1 TWINVEST DATA MANAGEMENT REGISTER TOOL STRUCTURE	11
--	----



LIST OF ABBREVIATIONS

C	Mechanical Designs
CL	Contact List
DEM	Hardware Prototypes
DMP	Data Management Plan
DOI	Digital Owner Identifier
EC	European Commission
ED	Experimental Data
FAIR	Findable, Accessible, Interoperable and Reusable
M	Models
ME	Marketing and Exploitation data
NDA	Non-Disclosure Agreement
S	Electronic Designs
SW	Software Outputs
TWINVEST	Universal, open-source and cybersecure Digital Twin to provide investors in onshore wind farms valuable insights about current operations and future investments
WP	Work Package



1 EXECUTIVE SUMMARY

The European Union's Horizon Europe research and innovation program, under Grant Agreement number 101146936, funds the TWINVEST project, which outlines its foundational data management plan (DMP) in this deliverable. The principal purpose of the DMP is to furnish a thorough overview of all datasets gathered, generated, and distributed during the project.

It also describes the data management techniques used for these datasets by the TWINVEST consortium. The original version of the DMP outlines the methods and standards employed, along with the status of the data, such as its generation, processing, or collection. It also talks about the Consortium's methods for data curation and preservation, as well as the possibilities for data sharing and transparency.

A thorough framework and technique for managing data inside TWINVEST are established by the preliminary version of the Data Management Plan (DMP). The framework and technique encompass both the administrative and technological aspects of data management. According to the FAIR principles—findable, accessible, interoperable, and reusable—these guidelines regulate data management. We intend to update the DMP on a frequent basis as the project develops. The TWINVEST project will improve the policy elements and provide more thorough information about the collected and created datasets by putting these modifications into practice.

2 OBJECTIVES

This deliverable will outline a comprehensive plan for the management of research data. We will implement this plan throughout the entire TWINVEST project and even after its conclusion. Even after the project is finished, it will clarify the procedures for data care and conservation, as well as the types of data that will be gathered, processed, and produced. It will also decide whether these data are publicly available. This document contains the initial Data Management Plan, but it will undergo consistent revisions over the project's duration, culminating in a finalized version by the project's completion.

3 INTRODUCTION

Internal recommendations for improving and optimizing access to the massive amount of data expected from all project segments and tasks are provided by the TWINVEST Data Management Plan. The "Task 9.3 Data Management Plan" is the cause of D9.3. Hence, to guarantee that all research data is readily discoverable, accessible, interoperable, and reusable, it will be amended in D10.1 "Final Data Management Plan" (M42).

4 DATA MANAGEMENT PLAN



This section provides an overview of the TWINVEST Data Management Register Tool, focusing on its functions in data gathering, documentation, and upkeep. It has a rulebook for new IP data, a registry for continuous updates, and a taxonomy of expected data types. To guarantee efficient and secure data administration that can adapt to the project's evolving demands, the plan also incorporates a metadata management architecture that handles data reference, ownership, contributors, and secrecy.

4.1 Data Management Plan on TWINVEST

Here, we review the TWINVEST steps to gain a comprehensive understanding of all the data generated throughout the project. To clarify, we can view the current status of all the data in the Data Management Register under "RESULTS" in Table 1, which is labelled as the "TWINVEST Data Management Register Tool."

4.1.1 Data Management Methodology

- a) All data generated during the project will be registered and documented in the "Data Management Register tool" of TWINVEST, which is hosted on the project's SharePoint, to ensure orderly administration of data sets. You can think of this register as a living document. Partners in the project will be responsible for regularly updating and filling it out. You can access this registry online through the TWINVEST SharePoint site. Following this, you will see a graphic depicting the Data Management Register Tool.
- b) The coordinator (RSTER) and any partners involved in its creation (if any) must be notified before a new Intellectual Property Data Item is logged.
- c) Subsequently, the data owner may input the updated IP-Data into the Data Management Register Tool. The following sections outline the required metadata for entering a new IP Data item.

4.1.1.1 Data type and how to reference

We predict that TWINVEST will generate a wide variety of data kinds, and we will use the Data Management Register Tool to collect these data types. We currently recognize the following data types, along with the corresponding shorthand notations. As the project advances, additional data types may become available, necessitating the creation of new acronyms.

Currently Acknowledged Data Types (together with their abbreviations):

- Experimental data (ED) originates from laboratory studies of all kinds. The characteristics of the experiment, along with the specific data recording tools employed by each laboratory, influence both the structure and the amount of data generated.
- Models (M) encompass algorithms, simulations, scripts, or code, all of which depend on the specific development platforms utilized.
- Design-related information, encompassing:



- Electronic designs (S) consist of schematic diagrams, block representations, material lists, and related components.
- Mechanical designs (C): 3D renderings, CAD drafts, etc.
- Hardware prototypes (DEM) will have their digital representations preserved via descriptive metadata and accompanying documentation, which will be made available in the results section of the project website.
- Software outputs (SW): Programs developed under the project's framework.
- Contact lists (CL): We utilize external contact lists to facilitate dissemination and communication with stakeholders.
- Marketing and exploitation data (ME): Data collected and processed for marketing and exploitation purposes.

The aforementioned data classifications do not include other document categories like deliverables or academic articles. The project Grant Agreement outlines distinct references and IDs for the former, while dissemination initiatives assign a Digital Owner Identifier (DOI) to the latter.

4.1.1.2 Metadata in Data Management Register Tool

1. Reference: We will designate a specific reference for each dataset based on its data category, Work Package and Task, data sequence, and sub-data sequence.

Data Number and Sub number

Data Type:

- WP, Task (and any subtask, if applicable): We will link each type of data to the specific WP, task (and any subtask, if applicable) from which it originates.
- Data Number: This denotes the quantity of datasets of a particular data type generated within a specific WP and task, presented in a sequential manner. All datasets of the same type within a task should have the same data number; use the sub-data number (described below) to identify discrepancies.
- Data Subnumber: This identifies datasets from a specific Work Package and task that are part of the same category and share the same Data Number. It is only identified once the data is created.

Examples of References:

DataType_WP.Task(.Subtask)_DataNumber_DataSubnumber.

- M_5.1_1 (Models, generated in WP 5. Task 1, from weather forecasting, short-range forecast).
- M_5.1_2 (Models, generated in WP 5. Task 1, from weather forecasting, mid-range forecasting).
- M_5.1_3 (Models, generated in WP 5. Task 1, from weather forecasting, seasonal forecasting).

2. Title: The name given to each IP Data Entry.



3. Data proprietor: The IP Data Entry's primary holder is identified. Joint ownership is only permissible when properly clarified in Column 6 (Dataset Origin) and Column 9 (Data Application) in Table 1. Shared ownership of IP data requires an NDA (non-disclosure agreement) at the end of the project.

4. Contributing Partner: All those who played a role in crafting the IP Data Item.

5. Creation Date: The project month when the data is anticipated to be produced.

6. Dataset Origin: Pertains to the Partner(s) supplying the data essential for the IP Data Item and/or the WP where it originates.

7. Purpose: Reason for the IP Data item.

8. Format: Physical for DEM items. For digital entries, it varies based on the software utilized.

9. Data Application: Relevant only during the project's duration.

10. Confidentiality: CONFIDENTIAL/SENSITIVE (SEN data treats by FAIR data principles).

4.1.2 Data collection

Each project's data collection process is unique due to factors such as the type of data generated, the informational requirements of project partners, and the methods used to store, retrieve, and protect the data. Therefore, the TWINVEST project's Data Management Register Tool necessitates the meticulous collection of all dataset's pertinent metadata in Table 1. We must perform this task manually. As stated in the "Data Management Plan" deliverable (D9.3), we will use this Excel file to gather data for the TWINVEST project.

4.2 FAIR DATA

This section describes the implementation of FAIR data principles in the TWINVEST project. FAIR stands for Findable, Accessible, Interoperable, and Reusable, in accordance with the [GO-FAIR guidelines](#). The main aim of these principles is to enhance the global utilization of SENSITIVE (NON-CONFIDENTIAL) data, promoting more efficient and widespread application of technical and scientific achievements. This approach aims to optimize resource utilization, thus avoiding redundant economic expenditures.

4.2.1 Making Data Findable, including provisions for Metadata

The development of comprehensive metadata, which improves the findability and discoverability of data, is the cornerstone of obtaining FAIR data inside TWINVEST. This is because metadata boosts the ability to find and discover data. Consequently, the project may require the addition of additional fields to the Data Management Register Tool at some point. Utilizing the "Reference" column makes it easier



to identify the data under consideration. The reference is necessary to ensure the uniqueness of each individual piece of data.

The project's SharePoint already has several pre-defined fields for different kinds of data, such as deliverables and publications. There will be a SharePoint area for the data register tool, much like there is for project deliverables and publications. Its structure allows project partners to review and modify it as needed. Also, there's a way to add new fields to TWINVEST, so it can handle more data in the future. This technique guarantees scalability and versatility in data management.

4.2.2 Making Data Openly Accessible

The TWINVEST project invites all participating entities to use non-confidential and non-sensitive data in their dissemination efforts. We will provide data for public viewing on the TWINVEST website. However, we will thoroughly assess a significant portion of the project's sensitive data before making it public. According to the Grant Agreement's clause on "Open access to scientific publications," each participant must make scientific publications related to the outcomes of their projects publicly available.

The project mandates that all dissemination elements must include explicit acknowledgments. These include a disclaimer that releases the European Climate, Infrastructure and Environment Executive Agency (CINEA) from liability for the content of the document, acknowledgment of funding from the European Union's Horizon Europe Green Research and Innovation program (under grant agreement No 101146936), and the display of the EU emblem.

Finally, the "Dissemination Activities" database on SharePoint would comprehensively document all dissemination activities conducted throughout the project. This guarantees a thorough record of the project's outreach and communication activities.

4.2.3 Making Data Interoperable

The TWINVEST project must appropriately utilize metadata to achieve interoperability. We can achieve enhanced interoperability by implementing proper metadata standards, techniques, and naming conventions.

To promote and enhance the interdisciplinary interoperability of (meta)data sets, the project will use generally recognized ontologies and a globally applicable language, such as English. Given the ongoing growth of this process, the Data Management Plan (DMP) can incorporate any identified opportunities for improvement as needed. This method allows for the continuous improvement and modification of the project's data management strategies.



4.2.4 Increase Data Re-Use

Table 1 of the Data Management Register Tool for the TWINVEST project carefully assesses the reuse potential of each dataset. The "Confidentiality" classification of the dataset is crucial to this evaluation:

- We designate "public" datasets as reusable, while we deem "confidential" datasets unsuitable for reuse.
- Items with the label "Public" are recognized as reusable.

It is crucial to remember that the Consortium Agreement, which all TWINVEST participants must abide by, initially assumes the confidentiality of data created and shared during the research. However, we will make certain datasets reusable for distribution to reflect the collaborative nature of the project and the scholarly interests of some participants. The Data Management Register Tool's "Confidentiality" section plays a crucial role in classifying project-generated data for reuse.

4.3 Allocation of Resources

The direct person-month costs for the TWINVEST project include all activities related to data generation and handling, in line with the person-month allocations for each partner.

For data preservation purposes, SharePoint will be used for a period of up to two years following the project's completion. All partners will continue to have access to the data after this period through alternate means, as specified in the Consortium Agreement.

4.4 Data Security

The TWINVEST project is deeply committed to safeguarding its data, products, and services from unauthorized access or use. One aspect of this commitment is the implementation of comprehensive protective measures.

Data security is essentially the responsibility of the individual partners, who are responsible for ensuring its protection. These individuals bear the responsibility of safeguarding all shared, processed, and operational data within their premises. Only authorized workers within the project consortium will be able to access this data through a stringent access control system. Partners must use a secure method to transmit data when necessary. In some instances, this may involve the physical transit of data, as well as encrypted data channels, secure digital transfer mechanisms, or even physical transportation of data. We strongly recommend utilizing SharePoint for the aforementioned purposes.

Data that is saved on the web-based repository for the project, which is called SharePoint, is protected by security mechanisms that are automatically given by the platform. These mechanisms provide robust protection against unauthorized access, aligning with the current business requirements. There are security measures in place, including authentication protocols and firewalls, among other things.



4.5 Ethical and Legal Aspects

The scientific activities of the TWINVEST project should not raise any ethical issues.

From a legal perspective, all participating entities have agreed upon the terms of Article 4 of the TWINVEST Consortium Agreement, which governs the management and sharing of data within TWINVEST. This calls for careful data evaluation before sharing or making it available, in accordance with the communication and dissemination strategies developed for each dataset, to avoid any potential breaches of confidentiality.

It is important to note that TWINVEST's approach to ethical and legal matters may change depending on the project's scope and circumstances.

5 RESULTS

This section presents the tool's structure as detailed in Table 1, which reflects the expected outcomes derived from the different tasks executed in the project. The Data Management Register Tool notably excludes certain document types, including deliverables and scientific publications, from its list of data types. As the project advances, Table 1 will undergo systematic updates and revisions to ensure it precisely represents the current developments and modifications. Furthermore, each iteration of this deliverable will incorporate Table 1, now expanded to encompass all gathered information, providing a comprehensive and up-to-date summary of the data management within the project.

Table 1 TWINVEST Data Management Register Tool structure

Reference	Title	Data Proprietor	Contributing Partner	Creation Date	Dataset Origin	Purpose	Format	Data Application	Confidentiality

6 CONCLUSION

This early Data Management Plan, as detailed in the deliverable, outlines a preliminary plan for managing data produced in the TWINVEST project. The plan specifies the categories of data to be gathered, processed, and/or generated, the degree of public accessibility of this data, and the strategies for its curation and preservation beyond the project's duration.

The paper conforms to the European Commission (EC) requirements, striving to ensure that TWINVEST data complies with the FAIR principles—making it Findable, Accessible, Interoperable, and Reusable.



This page is dynamic and will receive occasional updates and enhancements. This approach will integrate novel concepts and papers arising from current conversations. We anticipate making revisions to this Data Management Plan during the project's duration, specifically in month 42.

7 REFERENCES

[1] Grant Agreement No. 101146936 – TWINVEST

[2] TWINVEST Consortium Agreement