Detailed Data Management Plan

Deliverable No: D1.2 v1
Work package: WP1

Official delivery date: 30.06.2024
Actual delivery date: 26.06.2024

Dissemination level: Public

Co-funded by the European Union

Project: 101136119 | HORIZON-CL5-2023-D3-01 | www.twineu.net

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. Neither the European Union nor the granting authority can be held responsible for them.

Document Authors
Juan Galeano (RWTH Aachen University) Ilias Zafeiropoulos (UBE) Esteban Pastor (ETRA)

Version	Date	Author(s)	Notes
0.1	25.03.2024	Juan Galeano	first draft
0.2	28.03.2024	Juan Galeano	Added IPR directory section (M3)
0.3	21.05.2024	Juan Galeano	Changed template and final updates before review
0.4	24.06.2024	Juan Galeano	Addressed review comments and extended the security measures in section 5

Responsible Partner	RWTH Aachen University
Checked by WP leader	Padraic McKeever (Fraunhofer), 25.06.2024
Verified by the appointed Reviewers	Dario Pellegrino (Engineering), 17.06.2024 Apostolos Kapetanios, (European Dynamics), 17.06.2024
Approved by Project Coordinator	Padraic McKeever (Fraunhofer), 26.06.2024

Table of Contents

List of Tables

List of Abbreviations and Acronyms

Acronym	Meaning
DMP	Data Management Plan
IPR	Intellectual Property Rights
FAIR	Findable, Accessible, Interoperable, and Reusable
DOI	Digital Object Identifiers
WP	Work Package
CSV	Comma Separated Values
XML	Extensible Markup Language
JSON	JavaScript Object Notation
URN	Uniform Resource Name
GDPR	General Data Protection Regulation
GA	Grant Agreement
CA	Consortium Agreement

Executive Summary

This document is the first version of TwinEU Data Management Plan (DMP). The current document contains an introduction to the topic of data management and references to best practices from other H2020 projects, which will be applied for the DMP. It is a living document that will be constantly updated and revised during the project duration. In a later version, this document will contain all information concerning the data handling of the TwinEU project, moreover, the document will also act as the Intellectual Property Rights (IPR) directory, containing the list of items of knowledge/intellectual property with explicit information regarding ownership, nature, status, dissemination, and protection measures.

1 Introduction

Effective data management is an important part of the TwinEU project. A key element of this is a clearly defined process for handling research data. This process should be transparent and accessible to ensure engagement from all stakeholders, especially because different partners will be involved in the identification of the data to be handled. The DMP describes TwinEU's approach to research data handling, providing answers to all important questions about data processing, including data security, licensing, origin of data, and format. The DMP describes the data management life cycle for the data to be collected, processed, and generated by the TwinEU project. The content of this document remains flexible and subject to change during the project's runtime, necessitating regular updates and revisions. This document is the initial version of TwinEU DMP, serving as a first draft that will undergo further refinement over the course of the project.

To make research data Findable, Accessible, Interoperable, and Reusable, following the FAIR principles, the DMP should include information on:

- the handling of research data during and after the end of the project
- what data will be collected, processed, and/or generated
- which methodology and standards will be applied
- whether data will be shared/made open access and
- how data will be curated and preserved (including after the end of the project).

In addition, this document will contain the list of items of knowledge that are identified during the duration of the project and will be used as an Intellectual Property Rights (IPR) live directory containing explicit information regarding ownership, nature, status, dissemination, and protection measures. In this regard, this document is expected to contain any established cross-licensing schemes for project components.

TwinEU DMP and the IPR directory are oriented following best practices identified from other H2020 projects [1][2] and material provided by the EU Commission [3][4][5][6]. Material from these sources was used throughout the following document.

1.1 Task 1.4

The development of the TwinEU Data Management Plan is performed in a dedicated Task 1.4. This task will continually assess and improve the requirements of knowledge exchange, ensuring research data and outputs are managed in line with the FAIR principles.

1.2 Objectives of the Work Reported in this Deliverable

The objective of this deliverable is first, the open science data management and second, the identification and management of the Intellectual Property Rights (IPR) by the TwinEU project.

1.3 Outline of the Deliverable

The FAIR data principles and the approach that TwinEU follows to accomplish them are outlined in Chapter 2.

Chapter 3 is about data summary and presents the template expected to be filled with data information, in a later version will contain the catalogue of data published by the TwinEU partners.

The allocation of resources for making the data “FAIR” is explained in Chapter 4.

Chapter 5 and Chapter 6 are dedicated to data security and ethical aspects defined in the project.

Chapter 7 acts as the IPR directory and will list all items of knowledge related to the project's work.

2 FAIR data

The FAIR data principle is required to be used in EU-Projects by the "Guidelines on FAIR Data Management in Horizon 2020" [5][5]. It should support the exchange of scientific data and lead to knowledge discovery and innovation. The FAIR data approach is described by the acronym:

- **F**indable data: The first step to make data reusable is to find them. Clear naming and versioning of (meta-) data, easy to find by both humans and computers.
- **A**ccessible data: Once the data have been found, clearly specify how the data can be accessible, including needed tools, protocols, authentication, and authorization.
- **I**nteroperable data: The published data uses standards and vocabularies that allow interoperability with applications and workflows for analysis, storage, and processing.
- **R**e-usable data: The goal of the FAIR is optimizing reusability; therefore, it is clearly defined when and for which duration data is made available and under which licensing the data was published.

2.1 Making data findable, including provisions for metadata

The FAIR Data Management Guideline asks:

- Are the data produced and/or used in the project discoverable with metadata, identifiable and locatable by means of a standard identification mechanism (e.g. persistent and unique identifiers such as Digital Object Identifiers)?
- What naming conventions do you follow?
- Will search keywords be provided that optimize possibilities for re-use?
- Do you provide clear version numbers?
- What metadata will be created? In case metadata standards do not exist in your discipline, please outline what type of metadata will be created and how.

The datasets identified in TwinEU will be published on the Zenodo repository [7]. Zenodo is a general-purpose open-access repository developed under the European OpenAIRE [3] programme and operated by CERN [10]. Zenodo is free to upload and free to access. It allows researchers to deposit research papers, data sets, research software, reports, and any other research related digital content. Moreover, an important characteristic of Zenodo is that it follows the FAIR principles.

The datasets will be produced and published by the TwinEU partners responsible for the Work Packages (WP) or demos producing the datasets and will be labeled to identify different attributes as it is shown in Table 2. Each dataset will contain a metadata file describing the resource and explaining the meaning of the data as described in Chapter 3.

Version numbers will be given to the datasets, to distinguish different versions of the dataset produced during the project.

Globally unique persistent identifiers called Digital Object Identifiers (DOI) will be given to the data resources once published in the Zenodo repository. Zenodo automatically registers a DOI for a record once you publish it. This identifier ensures that the record can be uniquely cited which is important for reproducibility and attribution of credit.

The datasets will be given appropriate keywords to make them easier to discover during search.

2.2 Making data openly accessible

The FAIR Data Management Guideline asks:

- Which data produced and/or used in the project will be made openly available as the default? If certain datasets cannot be shared (or need to be shared under restrictions), explain why, clearly separating legal and contractual reasons from voluntary restrictions.
- How will the data be made accessible (e.g. by deposition in a repository)?
- What methods or software tools are needed to access the data?
- Is documentation about the software needed to access the data included?
- Is it possible to include the relevant software (e.g. in open source code)?
- Where will the data and associated metadata, documentation and code be deposited? Preference should be given to certified repositories which support open access where possible.
- Have you explored appropriate arrangements with the identified repository?
- If there are restrictions on use, how will access be provided?
- Is there a need for a data access committee?
- Are there well described conditions for access (i.e. a machine readable license)?
- How will the identity of the person accessing the data be ascertained?

The first objective of this deliverable is the open science data management. Horizon follows the “Open Science” approach that focuses on spreading knowledge as soon as it is available using digital and collaborative technology. Partners are requested to make their scientific publications available as Open Access publications, and grant access to their data as open as possible and as closed as necessary.

TwinEU expects to publish its datasets as open data on the Zenodo repository. External entities, like researchers, can use the interface of the Zenodo repository to search and download the TwinEU

data. The Zenodo platform is freely available to be used by the public, and browsing or downloading the data can be done without the need to register or open an account. The data will be stored in standard formats to be freely accessible for all external entities to download.

2.3 Making data interoperable

The FAIR Data Management Guideline asks:

- Are the data produced in the project interoperable, that is allowing data exchange and re-use between researchers, institutions, organisations, countries, etc. (i.e. adhering to standards for formats, as much as possible compliant with available (open) software applications, and in particular facilitating re-combinations with different datasets from different origins)?
- What data and metadata vocabularies, standards or methodologies will you follow to make your data interoperable?
- Will you be using standard vocabularies for all data types present in your data set, to allow inter-disciplinary interoperability?
- In case it is unavoidable that you use uncommon or generate project specific ontologies or vocabularies, will you provide mappings to more commonly used ontologies?

Making data interoperable primarily relies on the implementation of suitable standards, data models and metadata, as well as an appropriate associated vocabulary (e.g., search keywords).

The data produced by TwinEU will be published with full explanations of the meaning of the data and its context in the accompanying metadata documentation. Using a text format for the data and providing full explanatory metadata will facilitate interoperability.

2.4 Increase data re-use (through clarifying licenses)

The FAIR Data Management Guideline asks:

- How will the data be licensed to permit the widest re-use possible?
- When will the data be made available for re-use? If an embargo is sought to give time to publish or seek patents, specify why and how long this will apply, bearing in mind that research data should be made available as soon as possible.
- Are the data produced and/or used in the project useable by third parties, in particular after the end of the project? If the re-use of some data is restricted, explain why.
- How long is it intended that the data remains re-usable?
- Are data quality assurance processes described?

TwinEU datasets are expected to be published under the Creative Commons CC-BY-SA 4.0 license [8]. This license allows the datasets to be used if the data source is accredited and if the same licensing conditions (CC-BY-SA 4.0) are applied to its derivative use. The datasets will continue to be accessible on Zenodo after the project.

3 Data summary

The FAIR Data Management Guideline asks:

- What is the purpose of the data collection/generation and its relation to the objectives of the project?
- What types and formats of data will the project generate/collect?
- Will you re-use any existing data and how?
- What is the origin of the data?
- What is the expected size of the data?
- To whom might it be useful ('data utility')?

This chapter is expected to be structured according to the different possible generators of data to separately define and describe their datasets. Categorization of data or scenarios is envisioned, to have a clean and structured catalogue. In the context of TwinEU, data used will be from different resources. These data types can include reports, surveys, images, network models, SCADA information, measurements, raw data from sensors, and spreadsheets among other possibilities. In terms of possible categories, it is expected to have data from: Grid, Measurements, Market, Forecast, Simulation results, Resource information, Use cases, and Pilots' KPI values. This initial list of categories will be updated accordingly during the project.

Open Science related to scientific publications, research data, and additional practices is covered in the Grant Agreement (GA), specifically in the section "Communication, Dissemination, Open Science and Visibility". In terms of research data management, the section specifies that it should be ensured open access to the deposited data under the latest available version of the Creative Commons Attribution International Public License (CC BY) or Creative Commons Public Domain Dedication (CC 0) or a license with equivalent rights.

Because all the datasets are expected to be published as open data, the privacy and security aspects will be largely common, as is the tool used for storage, see Table 1. This applies by default to all the datasets below. Only deviations from the default handling will be described on a per-dataset basis in the envisioned sub-chapters in this section. To create the catalogue and provide a summary of the datasets, Table 2 will be used by the TwinEU partners to identify different important attributes of the published data.

Table 1: Default Data Security, Privacy, and Storage for all Datasets

Factsheet	
Data security and privacy default handling	
Classification level of data	All datasets are open data, licensed by Creative Commons — Attribution 4.0 International — CC BY 4.0.
Data privacy	None of the datasets contains personal or private data.
Exploitation and dissemination	
Availability (long-term storage)	The datasets are available without any time limit on Zenodo https://www.zenodo.org/ under TwinEU. Search will be possible using the project number 101136119.

Table 2: Template to be used to describe datasets

Factsheet	
Data Category name	To be filled once there is a categorization.
Dataset name	The name of the dataset
Dataset description	A description of the dataset
Available at	URL
Source of the data	
Re-use of historical data	If it is new generated data or historical data.
Data from devices like live trial measurements, sensors	If the data is gathered from measurements and sensor readings during the course of the activities.
Origin of data	e.g., from an analysis, survey, methodology in a specific Deliverable.
Timeplan for dataset	Schedule for data collection, processing, and availability.
Format of the open datasets	
Format of the data	e.g., CSV, XML, JSON
Metadata and documentation	Provide the standard, data model, or vocabulary explaining the data. Some data models have documentation or a schema definition file, if your data schema has this, then provide the link or any existing identifier (e.g., URN).
Exploitation and dissemination	
Purpose of data collection/generation, relation to project objectives	Explains why the data is being collected or generated. This includes the specific aims and goals of the data collection process.

Data utility, usefulness to external parties	How the data could be useful and beneficial to stakeholders outside the project.

4 Allocation of resources

Each of the WP or demo leaders will be responsible for the preparation of the datasets and their storage on the Zenodo platform. They will each register on the Zenodo platform and will upload their own datasets. Moreover, it is expected for them to specify how frequently will they update Zenodo.

Long term preservation will be accomplished without extra cost using Zenodo.

5 Data security

Each partner is responsible for the storage, security, and recoverability of their own generated data (following their institution's practices). All developments carried out during the project will be designed to safeguard collected data against unauthorized use and to comply with all relevant national and international regulations. Best practices and advanced data security measures will be implemented, including compliance with data protection and privacy guidelines and principles, considering the EU General Data Protection Regulation (EU) 2016/679 (GDPR) [9].

Further development of concepts around the cybersecurity of the data is expected to be carried out in Task 3.4. The data exchange occurring in the project is expected to follow the Data Spaces specifications, which provide recommended building blocks regarding security requirements. In addition, data storage will fully comply with the national and EU legal and regulatory requirements.

Once identified, the datasets to be published should not contain any data that is considered sensitive and should be suitable for publication as open data.

Long-term preservation of the open datasets will be achieved by publishing the datasets on the Zenodo platform, which will provide secure storage of the datasets.

6 Ethical aspects

The reference for ethical and legal issues is the EU General Data Protection Regulation (EU) 2016/679 (GDPR). Ethics and values are covered in the Grant Agreement, Article 14, and Data Protection aspects like data processing can be found in Article 15. Task 3.4 will give special attention to issues surrounding data protection.

The datasets that will be defined in this report will be related to technical data and not personal data so that no ethical considerations will arise concerning the publication of these technical data as open data.

7 IPR directory

The task describes the importance of clearly assigning the ownership of the various pieces of software/hardware developed during the project and setting up cross-licensing schemes between partners to be able to reuse non-public parts of the project. As stated in the Grant Agreement, dissemination and use of knowledge generated in TwinEU is governed by the terms of the Grant Agreement and the Consortium Agreement (CA). The Consortium Agreement specifies the

management of the Project and the rights and obligations of the work between the Parties and determines the rules about results, background, and the access rights.

This section is intended to be used to list all knowledge items relating to the project's work (both background know-how and results developed in the project) and make explicit for each item its owner, nature, status, and dissemination and protection measures. The background for the Project was identified and agreed upon by the Parties in the Consortium Agreement, as described in Section 9. In addition, the named section specifies the steps that need to be taken in case a Party wishes to modify or withdraw its Background. Finally, the section provides access rights specifications for background and results.

Results are described as achievements made during or shortly after the project's implementation and can be listed in the current section using the template as shown in Table 3.

For cross-licensing, a formal agreement is expected between parties and this section will contain specific parts of this agreement to provide useful insights and make it easy to find.

In addition, as stated in the GA, TwinEU partners will consider whether the project's outcomes can be exploited without infringing on existing patents claims and if its innovations can be patented.

Description of the expected information to be filled on the table:

- **Item name:** The title or name of the knowledge item or intellectual property.
- **Description:** A brief description of the item, including its purpose and significance. Add the WP or task related to the item.
- **Type of Ownership:** If the result has single or joint ownership.
- **Ownership [Country of establishment of the owner(s)]**: Who owns the intellectual property or knowledge and between brackets should be the country of establishment of the owner(s).
- **Nature**: The type or form of intellectual property (research findings, methodologies, software code, hardware, designs, etc.).
- **Status**: The current stage or status of the intellectual property (e.g., research phase, prototype development, finalized result, maturity level).
- **Related Documentation:** References to any documents or publications related to the item.
- **Contact Information:** Contact details for further information or inquiries about the item.
- **Dissemination Measures**: How the knowledge or intellectual property will be shared or disseminated.
- **Exploitation Measures:** Strategies and plans for utilizing the item, such as marketing strategies, partnership opportunities, or integration into products or services.

- **Protection Measures**: Any steps taken or planned for protecting the intellectual property (such as patents, copyrights, etc.). Details about the licensing terms (e.g., open-source, proprietary, Creative Commons).
- **Cross-licensing scheme:** refers to agreements between parties where each party grants the other party a license to use its intellectual property. Detailed as much as possible.

Table 3: Template to be used to describe results

Factsheet	
Item name	
Description	
Type of Ownership	
Ownership [Country of establishment of the owner(s)]	
Nature	
Status	
Related Documentation	
Contact information	
Dissemination Measures	
Exploitation Measures	
Protection Measures	
Cross-licensing scheme	

References

[1] "Data Management Plan" Deliverable 14.1 v3.0, Onenet H2020 EU Project, 2024. Available online: https://www.onenet-project.eu/wp-content/uploads/2024/04/OneNet_D14.1.pdf

[2] "Data Management Plan (final)" Deliverable 9.2 v1.0, Platone H2020 EU Project, 2021. Available online: https://www.platone-h2020.eu/data/deliverables/864300_M20_D9.2.pdf

[3] OpenAIRE Website. Available online: https://www.openaire.eu/

[4] Horizon 2020 Online Manual. Available online: https://ec.europa.eu/research/participants/docs/h2020-funding-guide/index_en.htm

[5] Guidelines on FAIR Data Management in Horizon 2020, European Commission, version 3.0, 2016. Available online: https://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf

[6] Your Guide to Intellectual Property Management in Horizon Europe, 2022. Available online: https://intellectual-property-helpdesk.ec.europa.eu/publications/your-guide-intellectual-property-management-horizon-europe_en

[7] Zenodo. Available online: https://zenodo.org/

[8] Creative Commons, "About CC Licenses". Available online: https://creativecommons.org/licenses/by-sa/4.0/

[9] Regulation (EU) 2016/679 of the European Parliament and of the Council, 27 April 2016. Available online: https://eur-lex.europa.eu/eli/reg/2016/679/oj

[10] CERN: European Council for Nuclear Research website. Available online: https://home.cern/