TWEET-IE
Twin Wind tunnels for Energy & the EnvironmenT
INNOVATIONS - EXCELLENCE

D6.3 Data Management Plan

LEAD BENEFICIARY: POLIMI

DISSEMINATION LEVEL: PUBLIC

M12 (31/10/2023)

Deliverable 6.3

WP6

Dissemination, Communication and Exploitation

Authors: A. Croce, V. Riziotis, D. Bouris, M. Manolesos, J. Prospathopoulos, P. Chasapogiannis, V. Pappa, N-P Pallas, K. Kellaris

TWEET-IE / Twin Wind tunnels for Energy and the EnvironmenT - Innovations and Excellence

HORIZON-WIDERA-2021-ACCESS-03-01 / PR# 101079125

Co-funded by the European Union

Contents

History of Changes

Ver	Date	Type	Changes	Contributors
1.0	04/04/2023	Report	Initial version	POLIMI, NTUA
2.1	31/10/2023	Report	2nd version, connection to the living document in the ARGOS platform	POLIMI, NTUA

Abstract

Deliverable D6.3 describes the Data Management Plan (DMP) of the TWEET-IE project. The aim of the DMP is to provide an analysis of the main elements of the data management policy that will be used throughout the project, with regard to all the datasets that will be generated and used. This deliverable is accompanied with a living document in the ARGOS platform [1], constantly updated as the implementation of the project progresses and when significant changes occur. The format of DMP follows the guidelines suggested for Horizon Europe calls.

Co-funded by the European Union

1. Data Summary

Results generated throughout the project are a priori destined to be reused as they will reinforce the capacities (technical, administrative etc) of the NTUA WT facility and will provide the basis on which future activities and networking can be based. Therefore, it is expected that the results generated throughout the project will be used for several validation studies in the future, at a national as well as international level.

The data types that will be generated during the project and their corresponding formats are listed in Table 1.

Data type	Formats
Experimental test data	Comma separated value (.csv), Excel (.xlsx), Structured plain text (.txt)
CFD simulation data	Comma separated value (.csv), Excel (.xlsx), Structured plain text (.txt), non-standard formats (e.g. paraview files)
Image files	Uncompressed tagged image (.tif, .tiff), JPEG (.jpg), Bitmap (.bmp), PNG (.png)
Video files	MPEG-4 (.mp4), AVI (.avi)
Documentation of experiments and simulations	PDF (.pdf), rich text (.rtf), Microsoft Word (.docx)
Deliverables, technical reports, publications	PDF (.pdf), Microsoft Word (.docx)

Table 1: Data types and corresponding formats

Partner	Long-term exploitation of data
Widening partner (NTUA)	Use the test / simulation data for blind tests and validations of CFD tools Share test / simulation data with institutions and industry Improve the reputation and research profile of NTUA WT facility and supporting staff
All	Promote the establishment of strategic networking activities between NTUA and top-class partners through training activities, WT network memberships, workshops, summer schools and related mobility of staff, young researchers, and doctoral students Collaboration with other universities, research institutes, private investigators and industries

Table 2: Long-term exploitation of data

Co-funded by the European Union

The expected testing results among WT research facilities are a valuable asset upon which all members of the consortium will draw for future research activities and innovative/creative proposals. Availability of such data is rare, therefore it is expected that other universities, research institutes, private investigators and industries will be highly interested in having access to the project results, also considering the fact that these will refer to three different topics: urban flows, wake interaction and micro devices for airfoil sections. The long- term benefits of the participants are summarized in Table 2.

Data generated from one setup of each WT test will be uploaded to the repository. Including the related photos and videos, a total size of 100GB can be considered sufficient for the storage of the total generated/re-usable data. Taking account that 4 twin tests will be performed, the analysis of the required data size is quoted in Table 3.

Data type	Data size per test case	No. of test cases	Total size
WT experimental	10 GB	4x2=8	80 GB
CFD simulations	2.5 GB	4x2=8	20 GB
Sum	12.5 GB	16	100 GB

Table 3: Data size analysis of generated / re-usable data

4 wind tunnel tests will be carried out at the facility of the widening partner (NTUA) and their 4 twins will be carried out at the facilities of the other partners. The generated WT test and simulation data will be originally hosted in the local storage space of each partner's facility. According to the procedure of each institute this could be a local repository or a file server or both.

2. FAIR data

2.1 Making data findable, including provisions for metadata

All datasets will be identified by a Digital Object Identifier (DOI) provided by the Zenodo repository service. In addition, rich metadata will be provided to allow discovery of the data. The types and content of metadata are summarized in Table 4 according to the FAIR principles.

Type of metadata	Content
Administrative	Name of project, funding agency (EU), names of contributors, project collaborator, project period, contact person
Descriptive	DOI identifier, name of the creator, i.e., name of the main researchers involved for producing the data, title, year of publication, subject describing the data and relative keywords in order to optimize the possibility for discovery and potential re-use
Structural	Resource (data) type (dataset, report, image, video etc.), data size and format (plain text, PDF, JPEG, MPG etc.)

Table 4: Type and content of metadata

All metadata will be indexed and researchable directly in the search engine of the Zenodo repository

Co-funded by the European Union

immediately after publishing. Furthermore, Zenodo's metadata will be transferred to the ARGOS platform through the creation of a living DMP document linked to the Zenodo's datasets. Metadata will be also transferred and indexed to the DataCite servers during DOI registration.

2.2 Making data accessible

All data will be deposited in the Zenodo trusted repository. Zenodo does not sign SLAs (service-level agreements) but is run by leading practitioners according to best practices. Zenodo ensures that data are assigned a globally unique and persistent identifier which is resolved to a digital object (DOI) to every published record. Accessibility and retrieval status of data is given in Table 5.

Type	Accessibility	Retrieval
Data	Open access, publicly available and findable in Zenodo for at least 10 years	Through DOI
Metadata	Open access, publicly available even when data are no longer available. Licensed under public domain dedication CC0.	Through record identifier and collection name using OAI-PMH [2] and REST [3] protocols. No authorization needed for retrieval.

Table 5: Accessibility of data and metadata. OAI-PMH [2] and REST [3] are open, free, and universal protocols for information retrieval on the web

No software will be needed to access or read the datasets. In case that a WT tunnel dataset is provided in a special format, the relevant software will be also uploaded in the repository. If data is present in a format that requires software for access, then provision will be made for the software to be open source (e.g., Paraview) with inclusion of a relevant access link.

2.3 Making data interoperable

Metadata vocabularies will follow FAIR principles. For example, the open external vocabularies of Open Definition (license), FundRef (funders) and OpenAIRE (grants) will be used. The JSON Schema [4] will be used as the language for the internal representation of metadata. JSON Schema offers export to other popular formats such as Dublin Core [5] or MARCXML [6]. Metadata will include qualified references to other metadata. Each referenced external piece of metadata will be qualified by a resolvable URL.

2.4 Increase data reuse

Increase of data reuse will be pursued through the following:

- Documentation files will be provided along with the datasets in order to give information on the measurement procedure or the simulation methodology and to explain the variables and the format of the datasets, thus facilitating data reuse.
- Data will be freely available in the public domain of the Zenodo repository to permit the widest reuse possible.
- Metadata will be licensed by Open Definition license.
- Repository data will be open to third parties during and after the end of the project.
- Data will be associated with detailed provenance.
- Metadata will refer the original contributors of the dataset.

Co-funded by the European Union

- All data and metadata uploaded will be trackable to any registered user (free registration to the repository).
- The selected repository allows curation and indexing of all relevant data per community harvesting. In this way, full text deposits will be linked to authoritative bibliographic metadata from third party systems, e.g., PubMed, Crossref, or SCOPUS where feasible.
- The widening partner (NTUA) will have the role of checking and curating all uploaded data.

Additional data produced during the project and stored by the individual partners will be available to third parties upon request. Requests may be placed to the project partners though the project web page (available for 10 years after the project) or the coordinator's web page (available indefinitely)

3. Other research outputs

Research outputs of the project will include WT test data, CFD simulation data, deliverables, technical reports and publications. The processed version of WT and simulation data will be uploaded to the selected repository and will conform to the FAIR principles as described in detail in the section 2. These data will be open to third parties and will be used for the blind tests and validation cases. Deliverables, technical reports, and publications will be also uploaded to the website of the project. Furthermore, research articles will be published to open access journals which use Digital Object Identifiers (DOI) and keywords to facilitate their findability.

Additional research outputs include all the raw WT and simulation data which are expected to have a large size and will only be of limited and specific interest to other parties. The status of those data regarding storage, accessibility and usage is described in Table 6.

Raw experimental and simulation data	
Storage	Separately by each partner hosting the experiment or performing the simulation in clouds or/and storage devices according to the policy of each institute
Accessibility	Full access to all partners of the consortium through the sharepoint of the project. Limited access to third parties upon request and according to the policy of the hosting institute. It can be made made through a private agreement between the host and the third party.
Reusability	Third parties will be informed for the existence of such data through the dissemination plan of the project (website of the project, social media, flyers, posters, exhibitions, presentations, publications). Long-term maintenance of the test and simulation data is foreseen from each partner ensuring their re-usability

Table 6: Status of raw experimental and simualtion data apart from the processed datasets uploaded to Zenodo repository

4. Allocation of resources

3 categories of data are distinguished with respect to the storage and maintenance resources: (a) The open access experimental and simulation data, (b) the large amount of raw experimental and simulation data (restricted access) and (c) the technical reports and publications. Table 7 summarizes the allocation resources for the storage and maintenance of each category.

Co-funded by the European Union

Output	Storage and maintenance	Cost
Open datasets	In Zenodo repository. Managed by engineers working for the widening partner (NTUA). No time limitation, maintained for at least 5 years according to Open Science practices	Freely provided by CERN. A personnel cost of 1kEuros is foreseen in WP6 of TWEET-IE project for uploading and updating the open datasets and the metadata
Raw data	By the hosting institute in sharepoint / cloud / other storage device for at least 10 years after the project	Personnel cost of 1kEuros per partner foreseen in WP6 of TWEETIE project plus in-kind contribution by project partners
Reports, publications	Web site of TWEET-IE project hosted on a central server of the widening partner (NTUA). Managed by engineers working for the widening partner (NTUA). Maintained for at least 5 years according to Open Science practices	1kEuros in personnel cost is foreseen for the construction and update of the website (WP6)

Table 7: Allocation of resources for data storage

5. Data security

The open/shared datasets will be securely stored, long term preserved and curated in the Zenodo trusted repository. The data published on the website will be also securely stored and preserved in the widening partner's server and external long term storage media. Each partner is responsible for the long-term preservation and curation of his own WT test and CFD simulation data. The suggested policy is to use a cloud (e.g., One Drive) and create a backup of all data in storage devices.

6. Ethics

As mentioned in Chapter 4 of the Description of Action (DoA), all activities carried out in the project are in compliance with the EU/national legal and ethical requirements of the countries where the tasks are to be carried out. There are no ethics or legal issues that can have an impact on data sharing and long-term preservation.

7. Other issues

The widening partner (NTUA) will make use of one of the Network Center's servers of NTUA in order to set-up, store and preserve the website of the project and its contents.

Co-funded by the European Union

References

1. ARGOS, “Plan and follow your data”, https://argos.openaire.eu/splash/
2. Marshall Breeding (September 2002). “Understanding the Protocol for Metadata Harvesting of the Open Archives Initiative”. Computers in Libraries 22 (8): 24–29.
3. Fielding, Roy Thomas (2000)."Chapter 5: Representational State Transfer (REST)".*Architectural Styles and the Design of Network-based Software Architectures* (Ph.D.). University of California, Irvine.
4. Open JS foundation, “JSON Schema”, https://json-schema.org.
5. DCMI, “DCMI Metadata Terms”, https://www.dublincore.org/specifications/dublin-core/dcmi-terms/
6. “MARC 21 XML Schema”. Library of Congress, https://www.loc.gov/standards/marcxml/

Co-funded by
the European Union