D6.3 Data Management Plan

LEAD BENEFICIARY: POLIMI

TYPE: DEC —Websites, patent filings, videos, etc

DISSEMINATION LEVEL: PUBLIC

Deliverable 6.3

WP6

Dissemination, Communication and Exploitation

Co-funded by
the European Union

History and Changes

Ver	Date	Description	Contributors
00	04/04/2023	Report	NTUA, POLIMI

Abstract

Deliverable D6.3 describes the Data Management Plan (DMP) of the project.

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 project will generate and re-use mainly experimental test and CFD simulation tabular data in comma separated values (.csv), Excel (.xlsx) or structured plain text (.txt) format. CFD output data may be also binary or ASCII files in a non-standard format, e.g., Paraview files. In addition, formatted documents in PDF (.pdf), rich text (.rtf) or Microsoft Word (.docx) format will be used for the description of the datasets. Image and video files will be also generated and uploaded to the repository in order to provide a representation of the wind tunnel tests. Images will be raster files in uncompressed tagged image (.tif or .tiff), JPEG (.jpeg), BMP (.bmp) or PNG (.png) format. Videos can be in MPEG-4 (.mp4) or AVI (.avi) format.

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. The top-class partners will benefit from the increased exposure and the broader area of application of their specific fields. Using the testing/simulation results for blind tests and validations of CFD tools and sharing them with institutions and industry, will improve the reputation and research profile of NTUA WT facility and supporting staff, which is one of the objectives of the project. Furthermore, the exploitation of the generated data through and beyond the duration of the project will reinforce the training activities, WT network memberships, workshops, summer schools and related mobility of staff, young researchers, and doctoral students. Therefore, it will promote the establishment of strategic networking activities between NTUA and top-class partners, which is another important objective of the project.

Data generated from one setup of each WT test will be uploaded to the repository. Including the related photos and videos, it is expected that the data size for one case will be ~ 10GB. Taking into account that 4 twin tests will be carried out during the project, the ensemble of 8 cases will generate a total of 80GB

data. On those, the corresponding CFD simulation data should be added. A total size of 100GB can be considered sufficient for the storage of the generated/re-usable data.

Four wind tunnel tests will be carried out at the facility of the widening partner (NTUA) and their four twins will be carried out at the facilities of the other partners. Therefore, 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.

The generated data will constitute a high quality and valuable research component in the form of benchmarking, validation, and novel research outputs. 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.

FAIR data

Making data findable, including provisions for metadata

Data will be identified by a Digital Object Identifier (DOI) provided by the repository service. Rich metadata will be provided to allow discovery of the data. Three types of metadata will be provided according to the FAIR principles: administrative, descriptive, and structural metadata. Administrative metadata will include the name of the project and the funding agency (EU), the names of the contributors, i.e., the project collaborators, the project period, and the contact person for managing the data. Descriptive metadata will include the DOI identifier of the data, the name of the creator, i.e., the name of the main researchers involved for producing the data, the title, the year of publication, a subject describing the data and the relative keywords in order to optimize the possibility for discovery and potential re-use. Structural metadata will include the resource (data) type (dataset, report, image, video etc.), the data size and the format (plain text, PDF, JPEG, MPG etc.).

All metadata will be indexed and researchable directly in the search engine of the repository immediately after publishing. Furthermore, metadata of each record will be sent to the DataCite servers during DOI registration and indexed there.

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.

During the project, a large amount of experimental WT as well as CFD simulation data will be produced. The consortium has agreed upon publishing the results from the twin test cases per WT test. These data will be publicly available to be used for blind test comparisons and validation of computational models. Any other data produced will be available to the partners of the consortium and its future use will be commonly decided on a case-by-case basis. Availability to external parties depends on the policy of each institute (host of the WT test) and can be made through a private agreement between the host and the third party. For those data that will be published through the repository there is no embargo period applied.

(Meta) data will be retrievable by their identifier using a standardized communications protocol. Metadata for individual records as well as record collections are harvestable using the OAI-PMH protocol by the record identifier and the collection name. Metadata will be also retrievable through the public

REST API. OAI-PMH [1] and REST [2] are open, free, and universal protocols for information retrieval on the web. Twin Test data will be publicly available without any restriction on the access during and after the end of the project. Metadata will be publicly accessible and licensed under public domain dedication CC0. No authorization will be necessary to retrieve it.

According to the FAIR principles, data will remain available and findable in the Zenodo repository for a period of at least 10 years after the end of the project. Metadata will be accessible, even when the data are no longer available. Metadata will be stored in the high-availability database servers of Zenodo, which are separate to the data itself.

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.

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 [3] will be used as the language for the internal representation of metadata. JSON Schema offers export to other popular formats such as Dublin Core [4] or MARCXML [5] Metadata will include qualified references to other metadata. Each referenced external piece of metadata will be qualified by a resolvable URL.

Increase data re-use

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 re-use. Data will be freely available in the public domain of the Zenodo repository to permit the widest re-use possible. Metadata will be licensed by Open Definition license. Data downloaded by the users will be subject to the license specified in the metadata by the uploader.

Repository data will be open to third parties during and after the end of the project. 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)

Data will be associated with detailed provenance. Metadata will also refer the original contributors of the dataset. 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. Furthermore, the widening partner (NTUA) will have the role of checking and curating all uploaded data.

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 sections 1 and 2. These data will be open to third parties and will be used for the blind tests and validation cases. The raw WT and simulation data are expected to have a large size and will only be of limited and specific interest to other parties. They will be stored separately by each partner hosting the experiment or performing

the simulation. Those data will be stored to clouds or/and storage devices according to the policy of each institute. Full access will be granted to all partners of the consortium through the sharepoint of the project and limited access will be granted to third parties upon request. Third parties will be informed for the existence of such data through the dissemination plan of the project which includes the website of the project, social media, flyers, posters, exhibitions, presentations, and publications. A long-term maintenance of the test and simulation data is foreseen from each partner ensuring their re-usability. Deliverables, technical reports, and publications will be 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.

Allocation of resources

The selected repository (Zenodo) is a free and open platform provided by CERN, fulfilling all FAIR principles, therefore there will be no cost for making the uploaded data Findable, Accessible, Interoperable and Re-usable. A planned personnel cost of the order of 1kEuros will be sufficient for the task of building the structure of the open datasets in the repository and furthermore for updating the data and the information. For the rest of research outputs, the large amount of WT test data and simulations will be stored by each partner separately, using a space in a cloud of his choice or/and some storage devices (e.g., hard disks etc.). Technical reports and publications will be uploaded to the web site of the project. The cost for the construction and update of the website is 1kEuros in personnel cost.

These costs will be covered from the budget of WP6 (Dissemination, Communication and Exploitation) of the project as well as in-kind contribution of project partners. The cost of 1kEuros per partner for storing data, as well as the cost of 1kEuros for the construction and update of the website will be allocated to partner personnel. The setup and update of the open/shared data in the repository, as well as the update of the website will be managed by engineers working for the widening partner (NTUA). Each partner will be responsible for managing the large amount data produced at his WT facility and the corresponding CFD simulations.

There is no time limitation on the preservation of the open/shared data in the Zenodo repository. They can be maintained there as long as CERN exists. In any case, complying with Open Science practices, they will be maintained for at least 5 years. The same applies for the preservation of reports and publications on the website which will be hosted on a central server of the widening partner (NTUA). Regarding the large amount of the WT data and simulations, the cost of 1kEuros per partner (mentioned above for the data storage) also covers the data preservation in sharepoints/ clouds/ storage devices for at least 10 years after the end of the project.

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.

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.

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.

References

1. Marshall Breeding (September 2002). "Understanding the Protocol for Metadata Harvesting of the Open Archives Initiative". Computers in Libraries 22 (8): 24–29.
2. 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.
3. Open JS foundation, "JSON Schema", https://json-schema.org.
4. DCMI, "DCMI Metadata Terms", https://www.dublincore.org/specifications/dublin-core/dcmi-terms/
5. "MARC 21 XML Schema". Library of Congress. Retrieved 2013-12-11, https://en.wikipedia.org/wiki/MARC_standards.