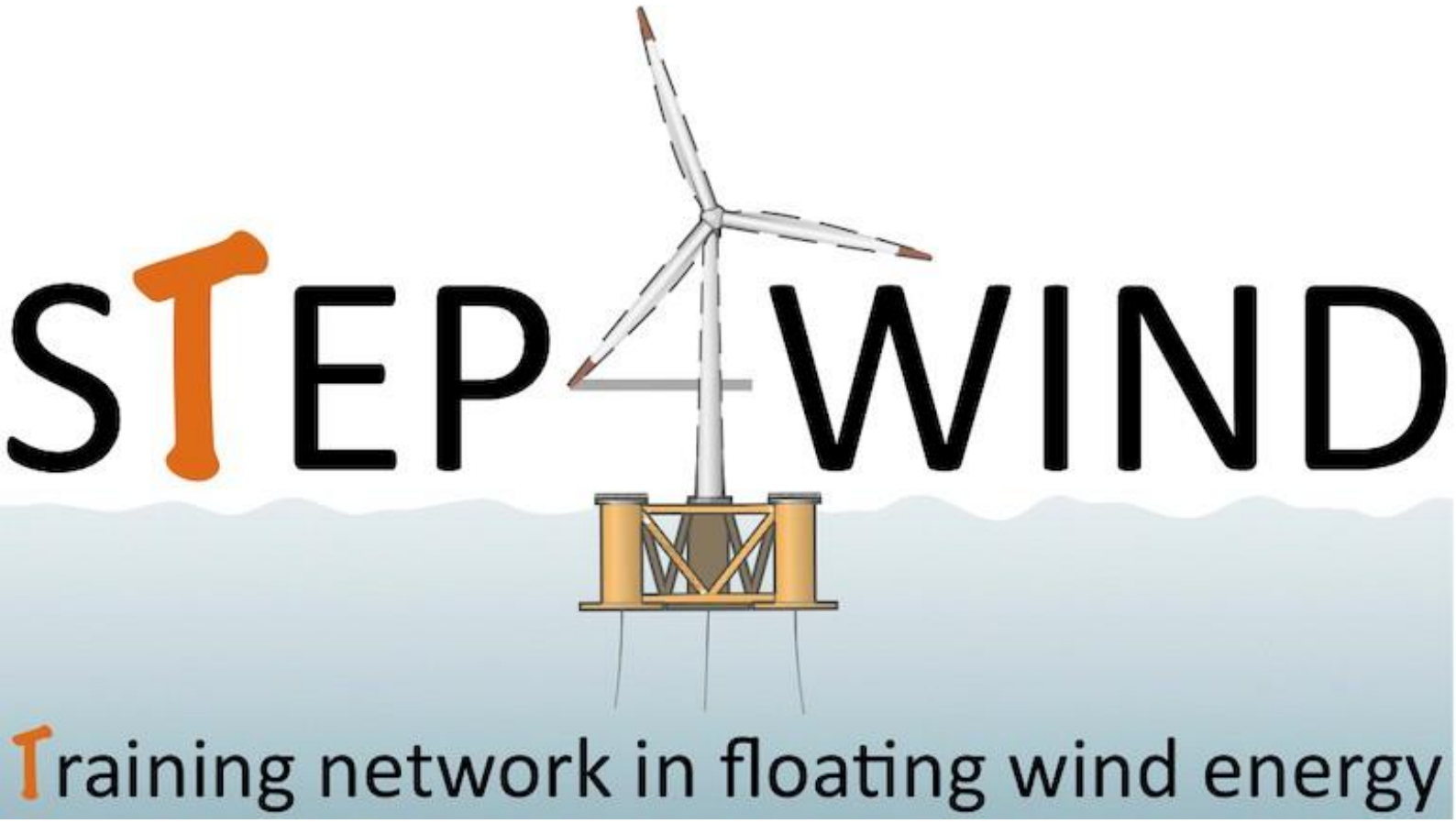
DATA MANAGEMENT PLAN

[D5.5, version 0.1]



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 860737.

Document History

Revision Nr	Author	Review	Date
0.1	Heather Andrews (TU Delft)	First draft	04.05.2020
0.2	Axelle Vire (TU Delft) Oana Schippers-Trifan (TUDelft)	Second draft (reviewed bij de Scientific Coordinator and partners)	25.05.2020
1.0	Heather Andrews (TU Delft)	Final version	30.09.2020

Table of Contents

List of Figures

1. Purpose of the document

1.1. Purpose of this document

A Data Management Plan is required for all projects participating in the extended ORD pilot. Thus, in STEP4WIND, the DMP describes the data management life cycle for the data to be collected, processed and/or generated its duration. As part of making research data findable, accessible, interoperable and re-usable (FAIR), the present document's structure is based on the EC defined template and addresses information on:

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

This DMP version is an initial defined plan within STEP4WIND with a final version to be submitted at the end of the project (April 2024); Nevertheless, the DMP is intended to be a living document in which information can be made available through updates as the implementation of the project progresses and when significant changes occur. Therefore, the current DMP version will updated in the context of the periodic evaluation/assessment of the project.

2. General Project Information

Title	Step4Wind:Novel de**S**ign, produc**T**ion and op**E**ration a**P**proaches for **f**loating **WIND** turbine farms
Acronym	STEP4WIND
Grant Agreement No.	860737
Funding Programme	REA - H2020-MSCA-ITN-2019
Instrument	Marie Skłodowska-Curie Innovative Training Networks
Topic	MSCA-ITN-2019 – EID Innovative Training Networks
Project Start Date	01/04/2020
Project Duration	48 months
Scientific Coordinator (TU Delft)	Dr. Ir. Axelle Viré Assistant Professor in Wind Energy TUDelft, Faculty of Aerospace Engineering Kluyverweg 1, 2629 HS Delft (NL) Email: A.C.Vire@tudelft.nl Phone: +31 15 27 81385 www.lr.tudelft.nl/windenergy
Project Coordinator (TU Delft)	Oana Schippers-Trifan TU Delft, Building 26 Valorisation Centre Van der Burghweg 1, 2628 CS Delft, The Netherlands Email: o.schippers-trifan@tudelft.nl Phone: +31 (0) 15 278 67 51

3. **Data summary**

Provide a summary of the data addressing the following issues:

- State the purpose of the data collection/generation
- Explain the relation to the objectives of the project
- Specify the types and formats of data generated/collected
- Specify if existing data is being re-used (if any)
- Specify the origin of the data
- State the expected size of the data (if known)
- Outline the data utility: to whom will it be useful

This project will manage different types of data:

1. Data/code related to research outputs (research results related to floating offshore wind turbines)
2. Data related to dissemination (e.g., related to outreach activities, the training of Early-Stage Researchers ESRs).
3. Data related to the management and progress of the entire project (project reports and deliverables)

Regarding the research output: there will be testing/training/validation data for machine learning algorithms, testing/validation data for the developed methodologies (including experimental data, numerical simulations and modelling data) and a final open source design analysis and optimization tool. During the project, data/in-house-developments of the respective partners will be accessible to the respective ESRs. The data/code generated will be managed locally by the respective partner in charge of such developments. The data/code will be stored in the respective institutional servers, managed by the respective IT department.

Main developments will be shared among the relevant partners via secure data/code transfer services provided by the leading partner in charge of such developments.

As research outcomes are finalized, all codes/models developed and the benchmark data used to derive such codes/models will be publicly released whenever possible (e.g., after checking for potential patentability). In addition to that, several dissemination outcomes will be delivered openly to the public based on the findings of the project (a MOOC over floating offshore wind energy, and an online game on designing floating wind turbines and farms). Training material derived from all the activities the ESRs will participate in might also be publicly shared to further increase the visibility and public engagement to the project.

Regarding the project management data, TU Delft (as the project coordinator) will have a separate restricted-access project drive within the university network, where all project reports, deliverables and main research outcomes will be stored. This project drive will be managed by TU Delft IT with automatic daily backups stored in different (physical) Data Centers. Remote access to the drive will be possible via SFTP protocol. Sharing of project documents among partners will be done via SURFdrive, SURFfilesender, email work addresses or via storing the data in the project drive set up by TU Delft.

In terms of data utility, the data/code that will be publicly released will be of use for the general public (to raise awareness and increase general knowledge about offshore wind energy) and to the scientific community (as new methodologies will be delivered for others to apply and built upon).

The overall size of the datasets is unknown at the moment. But it will be updated accordingly.

4. FAIR data

4.1. Making data findable including provisions for metadata:

The following diagram illustrates the FAIR data principles:

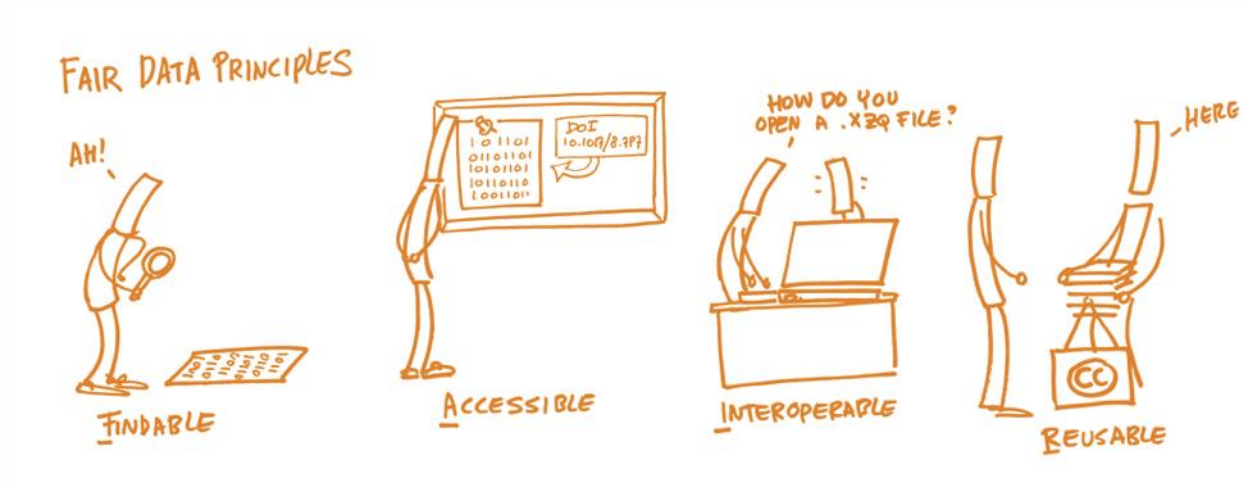


Figure 1. FAIR data principles (source: OPENAIRE)

- Outline the discoverability of data (metadata provision)
- Outline the identifiability of data and refer to standard identification mechanism. Do you make use of persistent and unique identifiers such as
- Digital Object Identifiers?
- Outline naming conventions used
- Outline the approach towards search keyword
- Outline the approach for clear versioning
- Specify standards for metadata creation (if any). If there are no standards in your discipline describe what metadata will be created and how

During the project, data/code under development will be safely stored locally, and appropriately shared among relevant team members (via secure data transfer services). Proper supporting and embedded documentation will be created as the datasets are generated. Version control will be used via the Git or SVN services provided by each respective leading partner (e.g., TU Delft Gitlab instance will be used for developments led by TU Delft). Each partner will be responsible for the management of the data/code during the project. Disciplinary metadata will be used during research to catalogue and index the data in a searchable manner. Such metadata will be clearly documented and used throughout the project. The idea is to have disciplinary metadata that is consistent and interoperable with what is already used in the field of wind energy. The documentation will then be openly shared alongside the data/code.

Data/code that can be open to the public will be archived via a trusted repository such as the4 TU.ResearchData or Zenodo. The new platform of the 4TU.ResearchData will be released in summer 2020, and it will provide integration with Git services (a service which is already provided by Zenodo). Thus in the near future (and certainly by the time the codes developed by this project are finalized) it will be possible to archive a snapshot of a public Github repository via the 4TU.ResearchData. This type of integration will be used to publicly archive the codes/models derived from this project.

The 4TU.ResearchData as well as Zenodo, both use by default standard citation metadata (Dublin Core) which makes the uploaded datasets easily findable by web search engines. Findability is further ensure by the fact that these services also provide persistent identifiers (DOIs) to the uploaded datasets.

Data/code that will have to remain under restricted access (e.g., commercially sensitive data) will be kept in a secure drive within the leading partner's institutional server, accessed only by relevant team members, and managed by the respective IT department. The data/code shall be stored in open/standard formats and with the proper documentation, including the disciplinary metadata convention (e.g., as an XML file, with documentation as a PDF file).

4.2. Making data openly accessible

- Specify which data will be made openly available? If some data is kept closed provide rationale for doing so
- Specify how the data will be made available
- Specify 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)?
- Specify where the data and associated metadata, documentation and code are deposited
- Specify how access will be provided in case there are any restrictions

As explained in the previous answers, all code will be openly released to the public via a e.g., Github public repository where a snapshot will be archived by a trusted repository (such as the 4TU.ResearchData or Zenodo). The codes will be written in open/standard programming languages. Supporting documentation will specify computational requirements and dependencies need to compile and run such codes.

Benchmark data will also be publicly archived via a trusted repository like the 4TU.ResearchData or Zenodo, unless there are commercial reasons not to disclose such data (e.g., data related to patentable inventions). The data will be released in open/standard formats, together with documentation specifying the necessary software/hardware/instrumentation used to generate and process such data.

A service like the 4TU.ResearchData or Zenodo ensures the findability and accessibility of the datasets in the long-term (> 10 years) by assigning a persistent identifier (DOI) to each uploaded dataset, and sharing it together with standard citation metadata. In addition to that, the data and code will be released under open licenses (open content licenses such as CC-BY in case of data, and open source license such as Apache in case of code).

Commercially sensitive data will be stored locally at a restricted-access drive managed by the IT department of the respective leading partner. The contact person for such data will be the assigned leader of the respective task. This person will be the one providing access to others whenever agreed among team members.

4.3. Making data interoperable

- Assess the interoperability of your data. Specify what data and metadata vocabularies, standards or methodologies you will follow to facilitate interoperability.
- Specify whether you will be using standard vocabulary for all data types present in your data set, to allow inter-disciplinary interoperability? If not, will you provide mapping to more commonly used ontologies?

Disciplinary metadata will be created accordingly. The idea is to follow the standards already used in the field of wind energy and built upon that when necessary. Datasets that will be publicly released via the 4TU.ResearchData or Zenodo will be archived with the citation **metadata standard** used by the chosen repository. Metadata values will follow **standard vocabulary.**

The codes will be developed in different programming languages such as FORTRAN, C++ and Python. They will be releasedo pen source (together with supporting documentation), so that it is easier for others to build upon or make interoperable with their respective developments. Input/output data used by the codes will also be in open/standard formats. The codes will also generate comprehensive log files to further facilitate the re-use.

4.4. Increase data re-use (through clarifying licenses):

- Specify how the data will be licenced to permit the widest reuse possible
- Specify when the data will be made available for re-use. If applicable, specify why and for what period a data embargo is needed
- Specify whether the data produced and/or used in the project is useable by third parties, in particular after the end of the project? If the re-use ofsome data is restricted, explain why
- Describe data quality assurance processes
- Specify the length of time for which the data will remain re-usable

All data/code that will be made publicly available will be released as **the respective journal articles are published** (in Open Access). For data/documentation, the datasets that will be publicly released via the 4TU.ResearchData or Zenodo will be published under an open content license, such as CC-BY. The codes shared publicly will be published under an open source license such as Apache. Both the 4TU.ResearchData and Zenodo ensure the datasets are findable and accessible for **at least 10 years.**

To facilitate further reuse of the codes, codes will be published in open/standard programming languages with comprehensive embedded documentation (e.g., function headers) and intuitive variables nomenclature. Supporting documentation will also be provided specifying processing power and hardware needed to run the codes, compilers, packages and libraries used, dependencies compatibility, etc. The aim is for others to build upon these codes and create a community around such developments.

Data that is commercially sensitive will be kept under restricted access until the data is not commercially sensitive anymore (e.g., patentable inventions). Such data will be managed according to the Legal advice available at the institutions and in agreement between all relevant partners.

environments, this will ensure that the ESRs are trained to the industry needs. It will also give the opportunity to the ESRs to acquire research or complementary training that would not have been possible at the academic or industrial institution alone.

Recruitment in STEP4WIND will be shared between two beneficiaries (see table 1.7 in the Grant Agreement).

5. **Allocation of resources**

- Explain the allocation of resources, addressing the following issues:
- Estimate the costs for making your data FAIR. Describe how you intend to cover these costs
- Clearly identify responsibilities for data management in your project
- Describe costs and potential value of long term preservation

Even though an estimation of the size of the publishable datasets is not yet known (but will be updated here accordingly), TU Delft researchers can upload up to1 TB of data free of charge per year to the 4TU.ResearchData. At Zenodo, researchers from other institutions can make uploads up to 50 GB per dataset. Overall those quotas are expected to be enough for the publishable outcomes. Hence no costs are foreseen regarding data/code publishing.

Data that will remain under restricted access will be stored in network drives of the respective institutions. This storage is already provided and available at the respective institutions (same applies for the storage of data/code during the project). Thus no costs are expected regarding data storage.

Regarding processing power the local processing power provided at the different facilities will be used. If further power is needed, it has been considered for TU Delft to apply to SURFsara supercomputing time.

When it comes to software, if licensed software is used, researchers will use the already available licensed software at the respective institutions. The outcomes of this project will be developed in open programming languages to encourage reusability.

Regarding responsibilities, each institution will be responsible for the data/code management related to the tasks/projects they are in charge of. TU Delft as the coordinator will be in charge of the management of project administrative data (including reports and deliverables). As mentioned before this data will be kept in a separate project drive within the TU Delft network, accessed only by relevant team members.

6. Data security

6.1. Address data recovery as well as secure storage and transfer of sensitive data

As mentioned above, the data/code generated during this project will be stored locally at the respective institution, in a secure restricted access drive within their local networks. The drives will be managed by the respective IT department. Remote access to such drives will be only via secure protocols.

TU Delft (as coordinator) will keep copies of all documents, reports and deliverables of the project. These will be stored in a secure restricted access project drive, backed-up on a daily basis and maintained by TU Delft IT. The drive will be accessible only to relevant team members via a secure protocol (SFTP).

For communication purposes related to research only work email addresses will be used. No auto forwarding of work emails to personal emails will be allowed.

For data transfer, SURFdrive or SURFfilesender (for Dutch researchers) or analogue services (for non-Dutch researchers) will be used among team members.

For any further assistance, the first point of contact for data/code management issues will be the Data Steward of the Faculty of Aerospace Engineering of the TU Delft.

7. **Ethical aspects**

To be covered in the context of the ethics review, ethics section of DoA and ethics deliverables. Include references and related technical aspects if not covered by the former

No ethical aspects are foreseen in this project. Collaborators have been carefully chosen to avoid the occurrence of conflicts of interest. Possible commercialization of data has been discussed among partners and will be properly addressed in the legal agreements of the consortium.

No personal data will be collected nor processed for research purposes. Videos or photos of dissemination activities (e.g., workshops, outreach) will be recorded after the consent of all involved participants. Such videos and photos will not be used for research purposes. They will only be used for dissemination purposes.

In any case, if any ethical questions arise, the respective collaborator(s) will seek advice from his/her(their) local data protection officer or ethics advisor. ESRs will also be encouraged to participate in Ethics-related courses or activities provided by the partner universities.

8. **Other**

Refer to other national/funder/sectorial/departmental procedures for data management that you are using (if any) TU Delft (as coordinator) will follow the TU Delft Research Data Framework Policy that establishes roles and responsibilities of researchers when it comes to research data (and code) management. Main principles of this framework are in line with those of H2020 regarding data management planning and FAIR principles adherence. And as such all team members will adhere to them.

References

- H2020 Online Manual
- H2020 Programme Guide for Applicants Marie Skłodowska-Curie Actions Individual Fellowships (IF). Version 1.4 / 08/04/2020