



Deliverable 8.3: Data Management Plan

WP8, Task 8.5

Date of document

30/06/2020 (M 6)

Deliverable Version:	D8.3, V.1.0
Dissemination Level:	PU/
Author(s):	<i>Mireia Olave, Raquel Hidalgo (IKERLAN)</i>

Document History

Project Acronym		iNTERESTING	
Project Title		Innovative Future-Proof Testing Methods for Reliable Critical Components in Wind Turbines	
Project Coordinator		Mireia Olave MOlave@ikerlan.es IKERLAN	
Project Duration		01/01/2020 – 31/12/2022 (36 Months)	
Deliverable No.		D8.3 Data Management Plan	
Diss. Level		Public (PU) /	
Deliverable Lead		IKERLAN	
Status		Working	
		Verified by other WPs	
	X	Final version	
Due date		30/06/2020	
Submission date		30/06/2020	
Work Package		WP 8 - Project coordination and management	
Work Package Lead		IKERLAN	
Contributing beneficiary(ies)		IKERLAN, VITO, BEC, LAULAGUN,KUL, VTT, MOVENTAS, SISW	
DoA		A Data Management Plan (DMP) will be issued on M6 and updated on a yearly basis. The DMP aims at making Research data findable, accessible, interoperable and reusable (FAIR) and will cover the handling, curation and preservation of data during and after the project, the type of data collected and generated by the project and the methodology and standards applied.	
Date	Version	Author	Comment
01/06/2020	0.1	Raquel Hidalgo	First draft of deliverable
15/06/2020	0.2	VITO, BEC, LAULAGUN,KUL, VTT, MOVENTAS, SISW	Inputs from Partners
30/06/2020	1.0	Raquel Hidalgo/Mireia Olave	Final draft

The content of this report represents the views of the author only and is his/her sole responsibility; it cannot be considered to reflect the views of the European Commission and/or the Executive Agency for Small and Medium-sized Enterprises (EASME) or any other body of the European Union. The European Commission and the Agency do not accept any responsibility for use that may be made of the information it contains

Table of Contents

1	Introduction	4
2	Data Summary	5
2.1	Purpose of data collection/generation	5
2.2	Data description	6
2.2.1	Types of data	6
2.2.2	Formats	6
2.2.3	Reuse of existing data	6
2.2.4	Size of data	6
2.2.5	Data utility	6
3	Fair data	8
3.1	Making data findable and interoperable	8
3.1.1	Standards and metadata	8
3.1.2	Search keywords	8
3.1.3	Use of persistent and unique identifiers	9
3.1.4	Naming conventions	9
3.1.5	Clear versioning	9
3.2	Making data accessible	9
3.2.1	Overview of access to data	9
3.2.2	Sharing and storing data among the consortium	10
3.2.3	Open access	10
3.3	Increase data reuse	11
4	Interoperability	13
5	Allocation of resources	14
5.1	Estimation of costs	14
5.2	Data management responsibilities	14
6	Data security & ethics	15
6.1	Protection of personal data	15
6.2	Confidentiality and integrity of data at rest	15
6.3	Sharing data with confidential access	16
6.4	Archiving confidential information	16
7	Bibliography	17

1 Introduction

ININTERESTING is a research & innovation project that aims to eliminate the need of building large test benches in the future by simplifying the product development process of new wind turbine components, reducing costs and time. To this end, a hybrid testing methodology that combines results from simplified physical tests and advanced virtual testing is proposed. This methodology will robustly predict reliability, lifetime and failures of full-scale wind turbine components through smart fusion process and upscaling techniques. Both the new turbine components as well as the hybrid testing methodology will be assessed by means of a comparative sustainability assessment. A comparative LCA (Life Cycle Assessment), LCC (Life Cycle Costs), and s-LCA (Social Life Cycle Assessment) will be carried out for the business-as-usual (BAU) wind turbine and the newly developed ININTERESTING solutions.

During the ININTERESTING project, data will be generated in a wide range of R&D activities. The ININTERESTING Data Management Plan (DMP) aims to provide a strategy for managing data generated and collected during the project, optimise the access to it and reuse of research data.

The DMP is intended to be a 'living' document that will outline how the ININTERESTING research data will be handled during and after the project. It describes the data management life cycle for all datasets to be collected, processed and/ or generated by the research project. It covers:

- Data handling during and after the project
- Data types and formats that will be generated/collected
- Methodologies and standards to be applied
- Whether the data will be shared or made open-access, and how
- How data will be curated and preserved

The present document is the 1st version of ININTERESTING DMP, containing a summary of the datasets; i.e. types, formats and sources (WPs and partner names) and specific conditions to be applied for sharing and reuse. This DMP will be reviewed and updated in months 18 and 36 of the ININTERESTING project.

This document follows the below structure in line with the EC template:

- Data summary
- Fair data
- Interoperability
- Allocation of resources
- Data security and ethics
- Bibliography

2 Data Summary

The DMP covers the complete research data cycle of iNTERESTING as described in Figure 1. In Step 1 of the DMP, iNTERESTING will produce raw data (generated through characterizations and simulations, surveys, ect). The data will then be processed and analysed into more usable forms; i.e. reports, publishable documents, data tables, codes.... In Step 2, the data will be preserved using appropriate naming rules and metadata schemes. The project's open access policy will be applied to determine which datasets shall be made accessible (share) for reuse in Step 3.

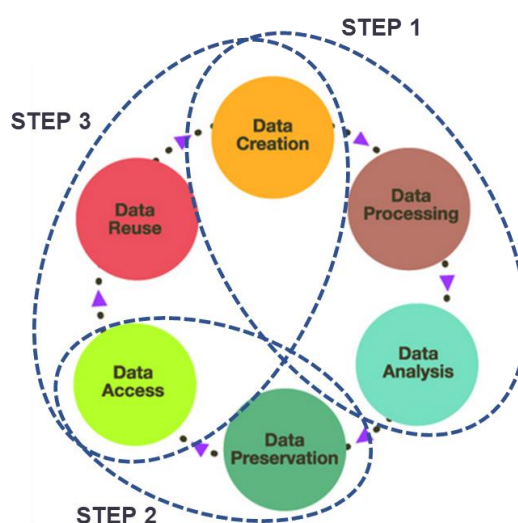


Figure 1. Research data life-cycle (Adapted from¹)

2.1 Purpose of data collection/generation

The main objective of iNTERESTING is to develop a hybrid testing methodology by combining results from simplified physical tests and advanced virtual testing through smart fusion process and upscaling techniques to robustly predict reliability, lifetime and failures of full-scale wind turbine components. iNTERESTING will generate and collect technical data in the frame of the following activities.

Table 1. Data generated in each of the workpackages/in the next activities of the iNTERESTING project

Activities	WPs
Development of new virtual tools for testing and design	WP2
Performance simplified experimental tests	WP3
Link and combine virtual and experimental results	WP4
Performance of real size tests	WP5
Assessment of the environmental (LCA), social (s-LCA) and economic (LCC) life cycle impacts and benefits of the newly developed solutions and testing method (WP6)	WP6

¹ <https://blogs.ntu.edu.sg/lib-datamanagement/data-lifecycle>

2.2 Data description

2.2.1 Types of data

The main data generated by the project are presented below:

- Data from materials characterization and failure tests
- Data from validation tests in real components
- Data representing the environmental (LCA), social (s-LCA) and economic (LCC) life cycle impacts and benefits of the newly developed solutions and testing method.

2.2.2 Formats

The formats of the data will include:

- Data and metadata
- MS Word (.doc, .docx) and Excel (.xls, .xlsx) compatible files
- PDF (.pdf)
- CAD (.iges, .step, .stl)
- ANSYS models (.wbpj)
- JPEG

2.2.3 Reuse of existing data

The reuse of existing data available from research projects and other European projects will be encouraged. In this sense, till now (M6), data from the following research projects was used as start point to develop the tasks of the iNTERESTING project:

- The CS1 (T2.1) is based on the model 20 MW RWT developed by Turaj Ashuri et al. ² (project INNWIND)
- The CS2 (T2.2) is based on the model DTU 10-MW RWT developed by Christian Baket al. ³ (project INNWIND)
- Two options are being considering as input data for the probabilistic load calculation methods (T2.1):
 - NEWA (New European Wind Atlas)⁴ which provides a unified high resolution and freely available data-set of wind energy resources in Europe.
 - LIDAR situated in the immediate vicinity of the virtual wind farm location.

This section will be updated at a later stage of the project.

2.2.4 Size of data

To be evaluated during the project and will depend on the extent and the nature of the data that are made available.

2.2.5 Data utility

The organizations that might be interested in the data generated during the project are:

- iNTERESTING consortium

² <https://github.com/tashuri/20MW-wind-turbine-model>

³ <https://www.hawc2.dk/Download/HAWC2-Model/DTU-10-MW-Reference-Wind-Turbine>

⁴ <https://www.neweuropeanwindatlas.eu/>

- Stakeholders involved in the field of Renewable Energy (manufacturers, remanufacturers, retailers, researchers) and stakeholders dealing with environmental concerns, socioeconomic and livelihood aspects.
- Scientific community
- European Commission services and European Agencies
- EU national bodies
- General audience



3 Fair data

3.1 Making data findable and interoperable

3.1.1 Standards and metadata

3.1.1.1 Datasets

In order to catalogue data in the repository as well as to facilitate their search and reuse, metadata will be filled in when uploading datasets in the Zenodo platform⁵ encompassing the following elements:

- Title
- Author
- Date
- Description
- Licence
- Funding

3.1.1.2 Publications

The publications issued during the project will include the Grant Number, acronym and a reference to the H2020 Programme funding, (Innovative Future-Proof Testing Methods for Reliable Critical Components in Wind Turbines-ININTERESTING-Nº 851245) including the following sentence:

“This project has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No 851245”.

When displayed together with another logo, the EU logo will have appropriate prominence.

Any dissemination of results must indicate that it reflects only the author's view and that the Commission is not responsible for any use that may be made of the information it contains.

Each paper must include the terms Horizon 2020, European Union (EU), the name of the action, acronym and the grant number, the publication date, the duration of embargo period (if applicable) and a persistent identifier (e.g. DOI).

The purpose of the requirement on metadata is to maximise the discoverability of publications and to ensure the acknowledgement of EU funding. Bibliographic data mining is more efficient than the mining of full-text versions. The inclusion of information relating to EU funding as part of the bibliographic metadata is necessary for adequate monitoring, production of statistics, and assessment of the impact of Horizon 2020.

3.1.2 Search keywords

When uploading the data and publications in Zenodo, data owners will enter keywords in order to facilitate the search. The list of keywords will be compiled during the project and included in later versions of the DMP.

⁵ Zenodo is a general-purpose open-access repository developed under the European OpenAIRE program and operated by CERN.

3.1.3 Use of persistent and unique identifiers

The assignment and management of persistent identifiers (PIDs) to the data will be automatically done when uploading the data in Zenodo repository.

3.1.4 Naming conventions

For metadata and dataset names, naming convention will be defined consisting in the following mandatory parts:

- A prefix, indicating if it is a dataset or metadata.
- A root composed by:
 - The short and meaningful name of the dataset
 - The acronym/short name of the data provider organisation(s)
 - A suffix indicating the date of the last upload into the Repository in YYYYMMDD format.

Each of these elements is separated by an underscore: _

3.1.5 Clear versioning

The files stored into the Repository will be named following the convention specified and the date will be placed as a suffix indicating the last version of the file uploaded into the Repository.

3.2 Making data accessible

3.2.1 Overview of access to data

The below table presents the different types of data, their accesses and the main guidelines for their management:

OA-Open access

CA-Confidential access

WP	Task	Type data	Access	Main exploitation guidelines
WP1	T1.1	Background data to be able to define the requirements for future bearings and gearboxes	OA	Open to public in final report
WP1	T1.2, T1.3, T1.4	Data from the case studies	CA	Confidential access for contractual reasons
WP2	T2.1	Data of probabilistic distribution of loads on the blade	OA	Open to public in final report/Zenodo
WP2	T2.3	Data of the induction hardening process	CA	Confidential access for contractual reasons
WP3	T3.1	Data from the characterization of the induction hardening process	OA	Open to public in final report/Zenodo
WP3	T3.2	Data from the characterization of the fatigue failure mode RCF	CA	Confidential access for contractual reasons
WP3	T3.3	Data from bonded joint properties.	OA	Open to public in final report/Zenodo
WP3	T3.4	Data from RSF characterization	OA	Open to public in final report/Zenodo

WP3	T3.5	Data from rolling element test results	CA	Confidential access for contractual reasons
WP3	T3.6	Multi-scale test data for journal bearing performance evaluation	CA	Confidential access for contractual reasons
WP3	T3.7	Test data for hybrid testing	CA	Confidential access for contractual reasons
WP5	T5.1	Data from RCF large test results and validation	CA	Confidential access for contractual reasons
WP5	T5.2, T5.3, T5.4	Data from large scale test for case studies	CA	Confidential access for contractual reasons
WP6	T6.1	Data to elaborate the sustainability assessment of BAU reference situation	CA	Confidential access for contractual reasons
WP6	T6.2	Data to elaborate the sustainability assessment of screening ININTERESTING solutions	CA	Confidential access for contractual reasons
WP6	T6.3	Data to elaborate the sustainability assessment results of ININTERESTING solutions	CA	Confidential access for contractual reasons

3.2.2 Sharing and storing data among the partners

To share and store data with the ININTERESTING consortium partners, a repository has been set up by Ikerlan. It provides access to data through a web interface and a platform to view, sync and share the files across devices easily — all under user's control.

3.2.3 Open access

Open access can be defined as the practice of providing on-line access to scientific information that is free of charge to the reader. In the context of Research and Development, Open Access typically focuses on access to “scientific information”, which refers to two main categories:

- Peer-reviewed scientific research articles (published in academic journals);
- Scientific research data (data underlying publications and/or raw data).

In line with H2020 Guidelines on Open Access to Scientific Publications, Open Access will be implemented in peer-review publications (scientific research articles published in academic journals), conference proceedings and workshop presentations carried out during and after the end of the project. Published articles (accepted author version) will be made public on the ININTERESTING website.

Two main routes exist for open access to scientific peer-reviewed publications:

- **Self-archiving** (also called “Green” Open Access) means that the published article or the final peer-reviewed manuscript is archived by the researcher – or a representative - in an online repository before, after or alongside its publication. Access to the article is often – but not necessarily - delayed (“embargo period”) as some scientific publishers may wish to recoup their investment by selling subscriptions and charging pay-per-download view fees during an exclusivity period. Depending on the journal selected, the publisher may require an embargo period between 6 and 24 months.
- **Open Access publishing** (also called “Gold” Open Access) means that an article is immediately provided in Open Access mode by the scientific publisher. The associated

costs are shifted away from readers and instead charged to (for example) the university or research institute to which the researcher is affiliated, or to the funding institutions supporting the research. When using this model, the costs of publishing are not assumed by readers and are paid by the authors, this means that these costs will be borne by the university or research institute to which the researcher is affiliated, or to the funding agency supporting the research. Gold Access will be opted for where possible.

Open Access will not affect the intellectual property generated by research results, as the decision on whether to publish Open Access documents will follow the procedure as described in article 8.4.1 of the Consortium Agreement to first seek protection for intellectual property rights as illustrated in the figure below:

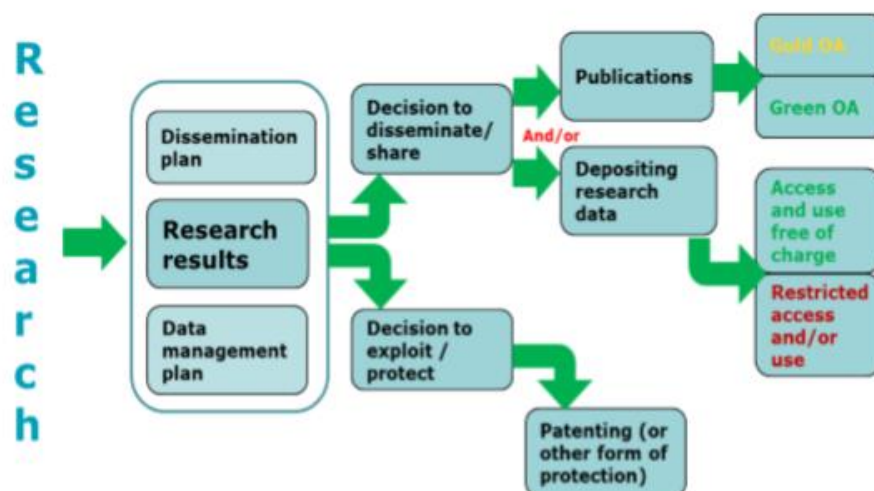


Figure 2. Open access to scientific publication and research data in the wider context of dissemination and exploitation⁶

The publications and data (except financial/contractual and personal data) related to the below list will be uploaded in Zenodo repository and will be given open access:

- Data of probabilistic distribution of loads on the blade
- Data from the characterization of the induction hardening process
- Data from bonded joint properties.
- Data from life extension tests
- Data from RSF characterization

The consortium will ensure that all publications issued by the ININTERESTING's partners are available as soon as possible, considering the embargo period (in case it exists).

3.3 Increase data reuse

This section will be compiled during the project as datasets will be made available during the for ININTERESTING project and shall include information on:

⁶ https://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-pilot-guide_en.pdf

Licensing of data

Availability of data and embargo period

Reuse of data by third parties

Data quality Insurance processes

Duration of data for reuse



4 Interoperability

This section will be compiled during the iNTERESTING project as datasets will be made available during the iNTERESTING project and shall include information on:

- Use of open software applications
- Use of standard vocabularies and methodologies



5 Allocation of resources

5.1 Estimation of costs

The establishment and maintenance of the repository created by Ikerlan for sharing the information of the ININTERESTING project (COLABORA) will not generate any costs. Additionally, uploading data to Zenodo repository will be free of charge.

Regarding publications in gold open access, this path shall be opted in where possible.

A full list of publications shall be annexed to the Data Management Plan during the project.

5.2 Data management responsibilities

Each ININTERESTING partner has to respect the policies set out in this DMP. Datasets have to be created, managed and stored appropriately and in line with applicable legislation.

Ikerlan, as WP8 leader, will ensure dataset integrity and compatibility for its use during the project lifetime by different partners.

Validation and registration of datasets and metadata are the responsibility of the partner that generates the data in the WP. Metadata constitutes an underlying definition or description of the datasets, and facilitate finding and working with particular instances of data.

Backing up data for sharing through open access repositories is the responsibility of the partner possessing the data.

If datasets are updated, the partner that possesses the data has the responsibility to manage the different versions and to make sure that the latest version is available in the case of publically available data.

Finally, all partners must consult the concerned partner(s) before publishing data in the open domain that can be associated to an exploitable result.



6 Data security & ethics

6.1 Protection of personal data

Foreseen personal data collection and processing, as well as protection measures, shall comply with the Regulation 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons about the processing of personal data and the free movement of such data ('GDPR') and national legislations.

The 'ethics requirements' that the ININTERESTING project must comply with are developed and included as a separate deliverable in WP9-Ethics.

6.2 Confidentiality and integrity of data at rest

Confidentiality and integrity of data at rest is covered by the ININTERESTING partners' data access policies and the security measures of the Zenodo Repository storage.

Partners's data access policies

- Allowing copies on local devices only during processing of the data with guaranteed erasure after being processed;
- extending the access control policies to the local copies;
- contractual clauses;
- agreement to terms and conditions before access is granted.

Additional security measures

Regarding the project collaborative area COLABORA set up by Ikerlan, this platform meets current safety regulations (ISO/IEC 27001 Information technology — Security techniques — Information security management systems — Requirements).

In addition, data at rest stored in the Zenodo repository will be protected against unauthorized access thanks to the following security measures/functionalities:

- The data centre is located on CERN premises and all physical access is restricted to a limited number of staffs with appropriate training and who have been granted access in line with their professional duties
- Servers are managed according to the CERN Security Baseline for Servers, meaning e.g. remote access to servers is restricted to Zenodo staff with appropriate training, and the operating system and installed applications are kept updated with latest security patches via our automatic configuration management system Puppet.
- CERN Security Team runs both host and network-based intrusion detection systems and monitors the traffic flow, pattern and contents into and out of CERN networks to detect attacks. All access to zenodo.org happens over HTTPS.
- Zenodo stores user passwords using strong cryptographic password hashing algorithms (currently PBKDF2+SHA512). Users' access tokens to GitHub and ORCID are stored encrypted and can only be decrypted with the application's secret key.

6.3 Sharing data with confidential access

When dealing with confidential data, partners will ensure that they comply with the non-disclosure policy detailed in section 10 of the Consortium Agreement. Thus, all information in whatever form or mode of communication, which is disclosed by a Party to any other Party in connection with the Project during its implementation and which has been explicitly marked as “confidential” at the time of disclosure, should be treated as Confidential Information.

6.4 Archiving confidential information

Partners collecting confidential data will archive them on secured internal servers and restrict their sharing in compliance with section 10 of the Consortium Agreement.



7 Bibliography

- European Commission. Guidelines on Fair Data Management in Horizon 2020. Version 3.0. 26 July 2016, http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-pilot-guide_en.pdf
- European Commission. Guidelines to the Rules on Open Access to Scientific Publications and Open Access to Research Data in Horizon 2020 Version 3.2. 21 March 2017, http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-pilot-guide_en.pdf
- European Commission. IPR Helpdesk. Fact Sheet Open Access to scientific publications and research data in Horizon 2020.
- Open access (Digital Agenda site): <http://ec.europa.eu/research/openscience/index.cfm>.
- European Commission. Grant Agreement number 851245 - iNTERESTING.
- iNTERESTING Consortium Agreement.



