With the contribution of the European Maritime
and Fisheries Fund of the European Union

DOCC-OFF

Project 863696 – DOCC-OFF – EMFF-BlueEconomy-2018

Scaling-up Digitalization Of Critical Components in OFFshore wind turbines

D1.2 Data Management Plan and IPR strategy

Data Management Plan (DMP): how data will be collected and processed within the project; what methodology will be adopted; how this data will be shared and/or made open; and how this data will be curated and preserved during and after the project to guarantee IPR protection and confidentiality.

DELIVERABLE INFORMATION

Deliverable title: Data Management Plan and IPR strategy

Due date: 30/06/2020

Date of submission: 30/06/2020

Work Package: WP 1 – Project Management and coordination

Dissemination level: Public

Deliverable leader: Basque Energy Cluster (BEC)

ABSTRACT

The present deliverable corresponds to **D1.2 Data Management Plan and IPR strategy.**

This deliverable is targeting a consistent management of data all along the DOCC-OFF data cycle by defining DOCC-OFF Data Management Plan (DMP) and Intellectual Property Rights (IPR). The DMP provides a broad analysis of the data that will be generated, processed and/or stored by DOCC-OFF partners. It provides a description of the methods to be used in terms of making DOCC-OFF data Findable, Accessible, Interoperable and Reusable (FAIR principle). The document also provides an explanation about the allocation of resources which includes the short/medium-term strategy and long-term strategy which assures DOCC-OFF generated data is preserved and accessible after the end of the project.

INDEX

INDEX OF FIGURES

EXECUTIVE SUMMARY

The EU-funded DOCC-OFF project main objective is to outline a condition monitoring strategy which may reduce the impact of some hydraulic pitch system failure modes on the wind turbine´s design load cases, and based on that strategy, to develop and validate a digital platform for component/system diagnosis. During the project partners will generate different types of data, which includes strategic documents, deliverables and reports, dissemination material, Scientific articles, aggregate data from the raw data of the analyzed pitch system and contest data to reproduce real pitch system operation scenario, together with experimental data from test sites that will be generated within WP2 and WP3.

This deliverable targets a consistent management of the data generated by defining the Data Management Plan (DMP). The DMP provides a comprehensive analysis of the data that will be generated, processed, stored and protected by the DOCC-OFF partners. The deliverable provides a description of the methods that will be used to make the DOCC-OFF data FAIR (Findable, Accessible, Interoperable and Reusable). The document also includes an explanation of the resource allocation including the short/medium term strategy and the long-term strategy that guarantees that the data generated by DOCC-OFF will be preserved and accessible after the end of the project.

1. INTRODUCTION

Data Management Plans (DMPs) are a key element of good data management. It describes the data management life cycle for the data to be collected, processed and/or generated by an EU-funded project. As part of making research data findable, accessible, interoperable and re-usable (FAIR), a DMP should include 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 apply
- Whether data will be shared/ made open access
- Which repositories will be used and,
- How data will be curated & preserved

This deliverable is a first version of the DOCC-OFF DMP, which will be updated over the course of the project whenever significant changes such as new data, changes in consortium policies, or changes in consortium composition and external factors arise.

Initially, this first version was scheduled to be produced during the first 6 months of the project, but due to changes in the consortium composition (Amendment AMD-863696-3 which replaces NEM Solutions for XABET) the action has been delayed until M8 (June 2020).

1.1 THE GENERAL DATA FRAMEWORK OF DOCC-OFF

The DOCC-OFF Data Management Plan (DMP) is designed with the idea of providing the necessary tools and mechanisms to promote adequate data management procedures.

The main objective of the project is to identify otherwise critical failure modes on the hydraulic pitch subsystem which may be made non-critical (or whose criticality may be reduced) via condition monitoring, and to develop a digital platform which, applied in the subsystem, will help to reduce the impact on the wind turbine´s design load cases (WP2). In a later stage, the digital platform will be performed in a real testing site in order to validate model simulation results (WP3). Most of the research data will be generated within these two work packages, which make up the bulk of the project. This data is set to be mostly confidential, only for members of the consortium (including the Commission Services), as specified in the Grant Agreement. However, it will be organized in such a way that all members of the consortium have access to them and can use them when necessary, to ensure quality of data processing at the monitoring and evaluation as well as other analytical work performed, and to help to the consortium to develop innovative services.

On the top of that, DOCC-OFF defines a set of dissemination and communications activities (WP4) that will heighten the project outputs by generating different material, publications, articles, e.g. This data will be completely public, and the process to handle and distribute it will be described within this document.

1.2 DOCC-OFF DATA CYCLE

The DOCC-OFF Data Cycle (Figure 1) covers the full project. It started with the signature of the Grant Agreement and the starting of the DOCC-OFF action. As soon as the DOCC-OFF consortium starts collecting data, it will be gathered and curated for preservation. Once the data is preserved, data owner will decide upon making data open or not. The decision on making open the research data may take place not immediately after the generation.

The DOCC-OFF Data Cycle includes both research data linked to confidential publications as well as any other digital data generated during the project.

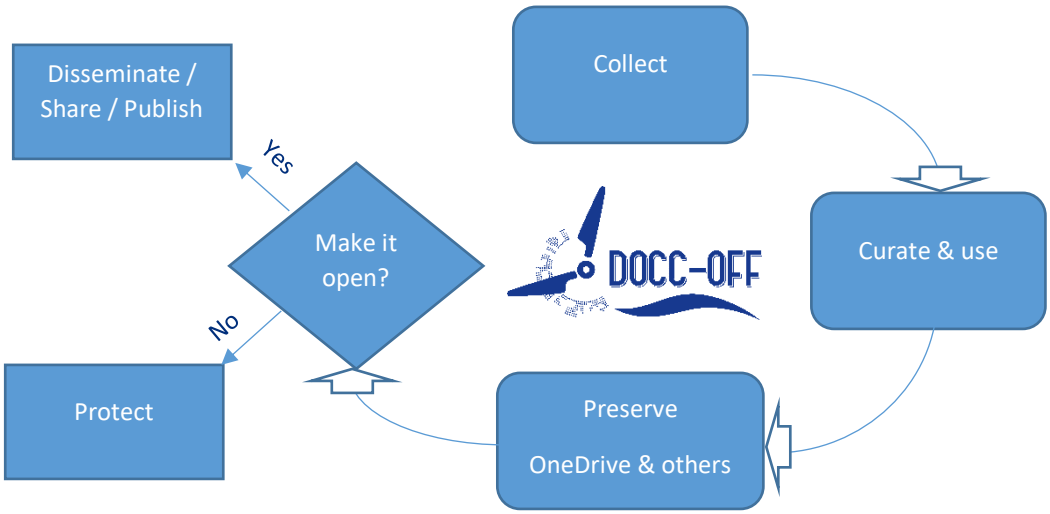

Figure 1. DOCC-OFF data cycle

2. DATA MANAGEMENT PLAN

2.1 RELATION TO THE OBJECTIVES OF THE PROJECT

The main challenge of DOCC-OFF is to outline a condition monitoring strategy which may reduce the impact of some hydraulic pitch system failure modes on the wind turbine´s design load cases, and based on that strategy, to develop and validate a digital platform that manages system operation data to know when and how the system fails in order to prevent this from happening.

For that purpose, it is necessary to develop a failure mode, effects and criticality analysis (FMECA) document considering offshore application, a hybrid model prototype to develop the algorithms and data analytics tools and a digital platform prototype which integrates the different technologies. The full integrated solution will be validated and demonstrated in a testing site to be defined.

2.2 SPECIFICATION OF THE ORIGIN AND TYPES OF DATA GENERATED AND/OR COLLECTED

Most of the data during the project lifespan of DOCC-OFF will be generated within the framework of the WP2, as a result of data processing for the correct development the digital platform that will capture and manage data from the pitch hydraulic system and will capitalize and exploit it through data analytics. It can be summarized as:

- Failure mode data and insights of the pitch system
- Raw data coming from physical signals of the pitch systems (temperatures, pressures, etc) and if possible, wind turbine environment such as wind speed, wind direction and other standard information.
- Plant metrics which will be calculated in the platform in order to understand the mean values of similar systems and assets.
- Aggregate variables from the raw data available in the project to enrich the analytical environment with new indicators such as relationship between physical variables.
- Rules and algorithms together with alarms generated by them which will suppose an event that will be recorded in the digital platform, as well.

Regarding digital platform validation and demonstration (WP3), data will be created directly from the performance of the integrated solution on a test bench under real conditions in an offshore environment, namely:

- The obtained datasets of the test bench parameters that monitor the settings and conditions of the real conditions that are reproduced on the test bench.
- The obtained datasets of the sensors implemented on the test-object.
- Thermal camera imaging data of the test set-up.

Additionally, any other digital data generated during the project includes:

- Strategic documents, deliverables and reports that will ease the replication and uptake of DOCC-OFF methods, technologies and solutions.

- Dissemination material in order to expose and exploit the results obtained during the project.
- Scientific articles or papers outlighting project progress and results.
- Pictures and video material of the test set-ups.

2.3 RESEARCH DATA

The raw data coming from pitch systems and wind information data of the farms will be collected as time series information and will be used for R&D of the project but could be of interest for future projects, too.

At the same time, over this raw data, new data will be generated thanks to the digital platform features. This new data will include mean values of different pitch systems, aggregated data as indicator or rules and even new parameters derived from algorithms to ensure the probability of future failure.

XABET will be responsible for the data management and privacy in those moments where confidentiality must be maintained on the digital platform. At the same time, all the aggregated data generated will be generated over the digital platform (cloud digital platform).

SIRRIS will be responsible to manage all related data linked to the validation testing of the test-object, this can include time series data from the test bench sensors, and test-object sensors but it can also include simulations results and/or models, pictures and video-images.

Data for the WP2 will be stored in a relational database in XABET´s cloud, following the IEC 61400 normative for renewables assets and projects. The stored data will follow the next policies:

- Encryption policies: all the communications between the digital platform that XABET will develop and the third parties will be encrypted and will follow an encryption policy.
- Password policies: all the user of the digital platform will have their own username and password with a one level authentication after a manual signed up process of every user that will be done by XABET. The password policy includes a strong procedure to ensure a secure password, allowing capital letters, numbers and more than 8 characters as password if is needed.
- Data processing policy: all the digital architecture behind the digital platform will be based on open source resources and without any kind of block box method. All the data through the platform will be processed and managed by transparent libraries and methods. XABET will host all the version control of this data processing procedure together with the code repository of the different modules and libraries that has been used for the digital platform development.

The digital platform will have a unique login gate for get into it. The described Data Policies are a core part of the data management practices. Moreover, the cloud storage will be done on the cloud with a recording and tracking of each login and navigation itinerary of each user through the platform and used data. Lastly, the digital architecture of the digital platform will generate automatic backups with an agreed frequency. This way the data protection and backup are guaranteed.

The data of WP3 will be stored on the SIRRIS Microsoft cloud (OneDrive) following the corporate data management policies within SIRRIS. Microsoft OneDrive will automatically have back-ups, access control and archiving.

2.4 FAIR DATA

All the data and documents created will follow FAIR principles. These do not necessarily suggest any specific technology, standard or implementation. Below, it is explained the solutions and options that DOCC-OFF will promote.

2.4.1 MAKING DATA FINDABLE

DOCC-OFF seeks to ensure that any data and supplementary material use standard formats and identifiers so that it is easy to find. In this sense:

- All reports, documents and deliverables will use clear versions following the structure established in the D1.1 Project Management handbook, so their status and evolution will be clearly recorded.
- DOCC-OFF encourages the use of a common language that will be used throughout the project.
- All presentations, contributions and peer review publications in particular will include the project logo and prominently acknowledge the grant agreement number and EU emblem.
- All the research data will be easily findable for the consortium members, following the data policies and tagging procedure explained before.
- An intelligent search tool will be available inside the digital platform to make even easier the data findable request.
- The IEC 61400 normative will be followed to name and storage the data, in order to follow a standard in the renewable sector

2.4.2 MAKING DATA ACCESSIBLE

DOCC-OFF project aims at making data as open as possible. In this sense, the default project public repository will be accessible from the project official website (https://www.doccoffproject.eu/en/) which will be constantly updated with latest documents and news.

Data will be accessible during the project for the associates through a web-based solution which will be the digital platform. Each user will have its own username and password and a proper URL and link to the digital platform will be hosted in the project's website. (https://www.doccoffproject.eu/en/)

In case of any particularities in sharing the data, these will be explained and justified.

2.4.3 MAKING DATA INTEROPERABLE

DOCC-OFF participants aim to generate interoperable data that enable the exchange and reuse of data. All systems will be well documented and, unless otherwise specified, open access. DOCC-OFF will follow established European metadata vocabularies, standards and methodologies.

There are three factors which ensure that all the data will be interoperable:

1. IEC 61400 normative will be followed to name and storage the data, in order to follow a standard in the renewable sector.
2. The digital platform will be developed over a standard API which will allow data downloading and integration if it is needed.
3. Data of testbeds is generally transformed to formats such as CSV or Excel formats in order to make sure they can be used by everyone; certain special formats can be extracted from these.

2.4.4 MAKING DATA REUSABLE

The possibility that a third person or entity makes use of a document entirely depends on the licensing conditions, as well as other intellectual property rights or permissions. The project website in this case is built with an “Open Source” licensed PHP, and the data is in a MariaDB database licensed "GNU General Public License".

The digital platform will be developed using a “postgre” database and using java as main programming language to ensure reusability. In fact, the open source libraries which will be used for the digital platform development will be enriched by XABET and all the development of interest over the open source libraries will be shared in the github (or similar) repository as a contribution from XABET.

In addition, the raw data and even the aggregate data generated in the platform will be available for downloading process if the user has rights to do it and it is not a confidential information.

3. DATA GOVERNANCE

The allocation of resources includes short-medium term and long-term strategies. The former defines who and how the expenses of making data fair will be covered as well as the data governance model.

The latter features the tools and methods that will allow the maintenance and preservation of data well after the project is finished. In this regard and in accordance with Article 13.1 of the Grant Agreement, *"the beneficiaries must - for a period of three years after the payment of the balance - keep records and other supporting documentation in order to prove the proper implementation of the action and the costs they declare as eligible".* Consequently, all digital material created within the tame frame of the project will be preserved for a period of 3 years after the completion of the project.

Confidential documents will be preserved in the OneDrive repository of the project, while public data will be preserved both in the OneDrive repository and on the website repository. The beneficiaries will keep the original documents, and they will keep them available upon request or in the contest of checks, reviews, audits or investigations.

CEPV as responsible for data management, will also save internally a copy of all documents generated in the project.

XABET will provide access to the data generated in WP2 once this access is suggested by one of the associates of the project and even with a third party if it is wanted or needed and proceeds. Data ownership will belong to the owner of the system which is generating the data and all the aggregated data generated in the digital platform will belong to the digital platform provider. The owner of the data will be responsible for implementing the data policies mentioned previously and for the preservation of the data and making it available in the long term. In order to reduce the cost of data preservation, XABET could hibernate the cloud of the project and a "wake up" process will be required to recover the data. This "wake up" process will be carried out by XABET.

Regarding the data generated in WP3, SIRRIS can provide access to the test data via the OneDrive platform if necessary. Data ownership will belong to the owner of the test-object in general. The data of the simulation models and data of certain conditions of the test bench are owned by SIRRIS.

3.1 DOCC OFF DATA GOVERNANCE MODEL

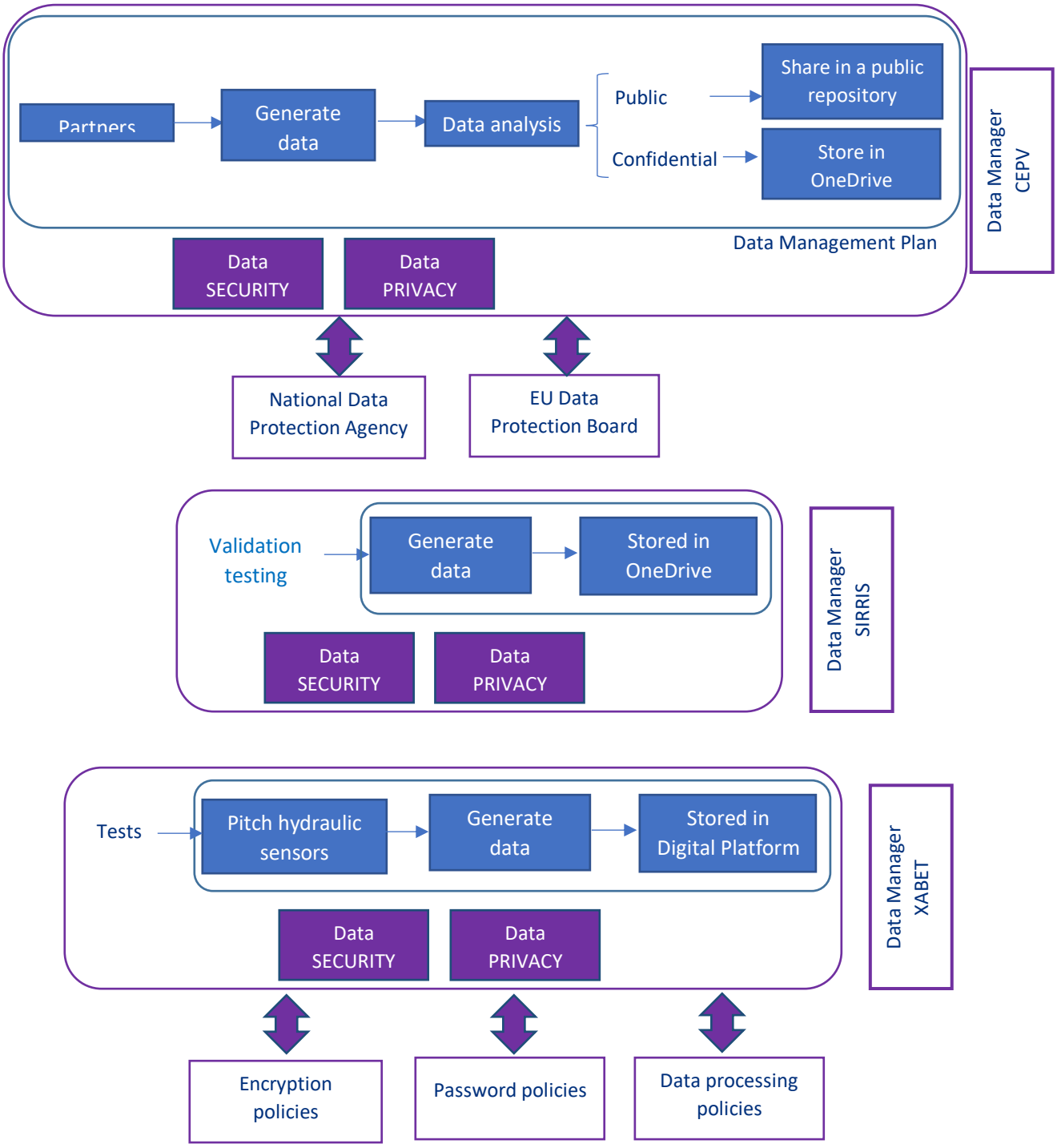

Figure 2. DOCC-OFF Data governance

4. INTELECTUAL PROPERTY RIGHTS (IPR) AND KNOWLEDGE MANAGEMENT STRATEGY

The overall strategy to ensure the proper IP and knowledge management and protection of DOCC-OFF will be set up in the Consortium Agreement to be signed by all partners. This way all project partners will agree on a legal framework and explicit rules to regulate ownership and access to key knowledge (IPR, data etc.), such as partner's already existing knowledge, rules to distribute new IPR created within the project, and protection measures to assure proper exploitation of innovative results.