Open Sea Operating Experience to Reduce Wave Energy Costs

Deliverable D8.6

Data Management Plan (update)

Lead Beneficiary	TECNALIA
Delivery date	2017-07-26
Dissemination level	Public
Status	Approved
Version	2.0
Keywords	Open access, datasets, metadata, ZENODO

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 654444

Disclaimer

This Deliverable reflects only the author's views and the Agency is not responsible for any use that may be made of the information contained therein

Document Information

Grant Agreement Number	654444
Project Acronym	OPERA
Work Package	WP 8
Task(s)	T8.2
Deliverable	D8.6
Title	Data Management Plan (update)
Author(s)	Pablo Ruiz-Minguela (TECNALIA), Endika Aldaiturriaga (OCEANTEC)
File Name	OPERA_D8.6_Data management plan_TECNALIA_20170726_v2.0.docx

Change Record

Revision	Date	Description	Reviewer
0.0	14-04-2016	Initial outline	Pablo Ruiz-Minguela
0.5	21-06-2016	Full draft contents	WP8 partners
0.9	15-07-2016	Version for peer review	Sara Armstrong (UCC)
1.0	27-07-2016	Final deliverable D8.5 to EC	EC
1.5	14-06-2017	Update of contents	WP8 partners
1.9	18-07-2017	Version for peer review	Sara Armstrong (UCC)
2.0	26-07-2017	Final deliverable D8.6 to EC	EC

EXECUTIVE SUMMARY

This document is an update of D8.5 Data Management Plan (DMP) for the OPERA project, which is funded by the European Union's H2020 Programme under Grant Agreement #654444. OPERA's main goal is to collect, analyse and share open sea operating data and experience to validate and de-risk four industrial innovations for wave energy opening the way to long term cost reduction of over 50%.

The Consortium strongly believes in the concepts of open science, and in the benefits that the European innovation ecosystem and economy can draw from allowing the reuse of data at a larger scale. Besides, there is a need to gather experience in open sea operating conditions, structural and power performance, and operating data in wave energy. Therefore, this project proposes to remove this roadblock by delivering for the first time, open access, high-quality open sea operating data to the wave energy development community.

The OPERA project participates in the Pilot on Open Research Data launched by the European Commission along with the H2020 programme. The use of a Data Management Plan is required for all participating projects.

The purpose of the Data Management Plan (DMP) is to provide an analysis of the main elements of the data management policy that will be used by the Consortium with regard to the project research data.

The DMP covers the complete research data life cycle. It describes the types of research data that will be generated or collected during the project, the standards that will be used, how the research data will be preserved and what parts of the datasets will be shared for verification or reuse. It also reflects the current state of the Consortium agreements on data management and is consistent with exploitation and IPR requirements.

Research data linked to exploitable results will not be put into the open domain if they compromise its commercialisation prospects or have inadequate protection, which is a H2020 obligation. The rest of research data will be deposited in an open access repository.

The DMP is not a fixed document; on the contrary, it will evolve during the lifespan of the project. This updated version of the DMP adds more details to the datasets and describes the practical data management procedures implemented by the OPERA project.

The expected types of research data that will be collected or generated along the project lie in the following categories: 1) Environmental monitoring; 2) Mooring performance; 3) Bi-radial performance; 4) Power output; 5) Power quality; and 5) Offshore operations.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABBREVIATIONS AND ACRONYMS

DMP	Data Management Plan
DoA	Description of Action
DOF	Degree of Freedom
Dx.y	Deliverable numbered, for example D8.6
EC	European Commission
KER1	Key Exploitable Result 1 (Floating OWC device and shared mooring configuration)
KER2	Key Exploitable Result 2 (PTO based on the biradial air turbine)
KER3	Key Exploitable Result 3 (Elastomeric tether for mooring systems)
KER4	Key Exploitable Result 4 (Advanced control algorithms for WECs)
KER5	Key Exploitable Result 5 (Open-sea testing and validation services)
KER6	Key Exploitable Result 6 (Guidelines and standards for ocean energy)
KER7	Key Exploitable Result 7 (Engineering and construction of wave energy farms)
KER8	Key Exploitable Result 8 (Seakeeping, installation and O&M services)
H_s	Significant wave height
T_p	Peak wave period
Tx.y	Project task numbered, for example T8.2
WP	Work Package

1. INTRODUCTION

1.1 OPERA MOTIVATION

The OPERA project participates in the Pilot on Open Research Data launched by the European Commission (EC) along with the H2020 programme. This pilot is part of the Open Access to Scientific Publications and Research Data programme in H2020. The goal of the programme is to foster access to research data generated in H2020 projects. The use of a Data Management Plan (DMP) is required for all projects participating in the Open Research Data Pilot.

Open access is defined as the practice of providing on-line access to scientific information that is free of charge to the reader and that is reusable. In the context of research and innovation, scientific information can refer to peer-reviewed scientific research articles or research data.

Research data refers to information, in particular facts or numbers collected to be examined and considered, and as a basis for reasoning, discussion, or calculation. In a research context, examples of data include statistics, results of experiments, measurements, observations resulting from fieldwork, survey results, interview recordings and images. The focus is on research data that is available in digital form.

The Consortium strongly believes in the concepts of open science, and in the benefits that the European innovation ecosystem and economy can draw from allowing the reuse of data at a larger scale.

Furthermore, there is a need to gather experience in open sea operating conditions, structural and power performance and operating data in wave energy. In fact, there has been very limited open sea experience in wave energy, which is essential in order to fully understand the challenges in device performance, survivability and reliability. The limited operating data and experience that currently exists are rarely shared, as testing is partly private-sponsored.

This project proposes to remove this roadblock by delivering for the first time, open access, high-quality open sea operating data to the wave energy development community.

Nevertheless, data sharing in the open domain can be restricted as a legitimate reason to protect results that can reasonably be expected to be commercially or industrially exploited [1] . Strategies to limit such restrictions will include anonymising or aggregating data, agreeing on a limited embargo period or publishing selected datasets.

1.2 PURPOSE OF THE DATA MANAGEMENT PLAN

The purpose of the DMP is to provide an analysis of the main elements of the data management policy that will be used by the Consortium with regard to the project research data.

The DMP covers the complete research data life cycle. It describes the types of research data that will be generated or collected during the project, the standards that will be used, how the research data will be preserved and what parts of the datasets will be shared for verification or reuse. It also reflects the current state of the Consortium agreements on data management and must be consistent with exploitation and IPR requirements.

FIGURE 1.1: RESEARCH DATA LIFE CYCLE (ADAPTED FROM UK DATA ARCHIVE [1])

The DMP is not a fixed document, but will evolve during the lifespan of the project, particularly whenever significant changes arise such as dataset updates or changes in Consortium policies.

This document is an update of the DMP which was delivered in Month 6 of the project (D8.5). It included an overview of the datasets to be produced by the project, and the specific conditions that are attached to them. The current version of the DMP gets into more detail and describes the practical data management procedures implemented by the OPERA project with reference with the IT tools developed in WP1. The final version of the DMP will be delivered in Month 30 (D8.7).

This document has been produced following the EC guidelines for project participating in this pilot and additional consideration described in ANNEX I: KEY PRINCIPLES FOR OPEN ACCESS TO RESEARCH DATA.

1.3 RESEARCH DATA TYPES IN OPERA

The data types that will be produced during the project are focused on the Description of the Action (DoA) and their results.

According to such consideration, Table 1.1 reports a list of categories of research data that OPERA will produce. These research data types have been mainly defined in WP1, including data structures, sampling and processing requirements, as well as relevant standards. This list may be adapted with the addition or removal of datasets in the next versions of the DMP to take into consideration the project developments. A detailed description of each dataset is given in the following sections of this document.

TABLE 1.1: OPERA TYPES OF DATA

#	Dataset category	Lead partner	Related WP(s)
1	Environmental monitoring	TECNALIA	WP1
2	Mooring performance	UNEXE	WP1, WP2, WP5
3	Bi-radial performance	IST	WP1, WP3
4	Power output	OCEANTEC	WP1, WP4, WP5
5	Power quality	UCC	WP1, WP5
6	Offshore operations	TECNALIA	WP6

Specific datasets may be associated to scientific publications (i.e. underlying data), public project reports and other raw data or curated data not directly attributable to a publication. Datasets can be both collected, unprocessed data as well as analysed, generated data. The policy for open access are summarised in the following picture.

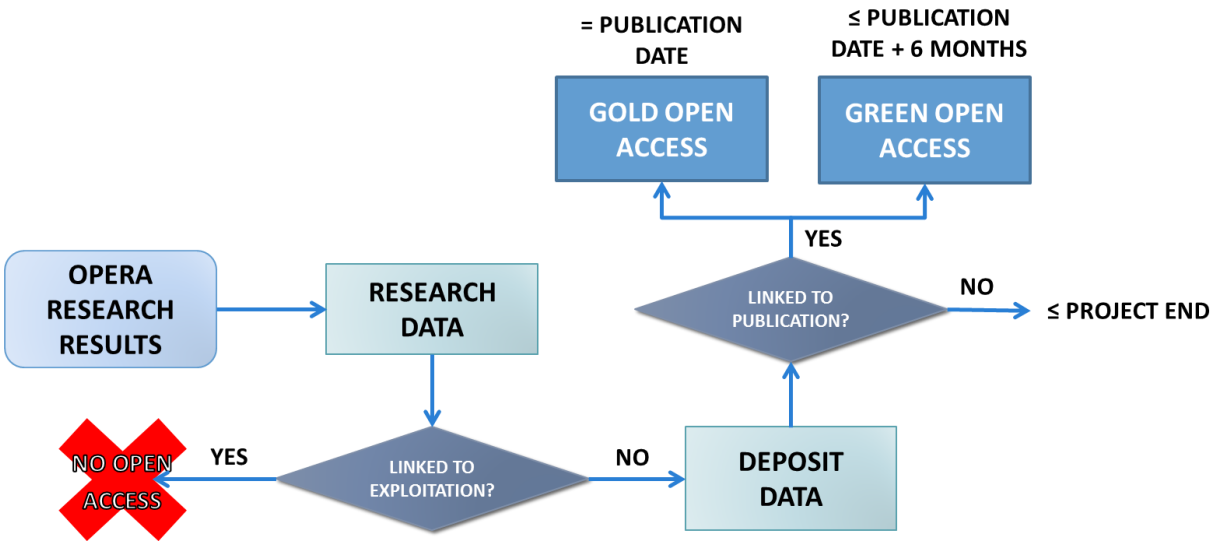

FIGURE 1.2: RESEARCH DATA OPTIONS AND TIMING

Research data linked to exploitable results will not be put into the open domain if they compromise its commercialisation prospects or have inadequate protection, which is a H2020 obligation. The rest of research data will be deposited in an open access repository.

When the research data is linked to a scientific publication, the provisions described in ANNEX II: SCIENTIFIC PUBLICATIONS will be followed. Research data needed to validate the results presented in the publication should be deposited at the same time for "Gold" Open Access[1] or before the end of the embargo period for "Green" Open Access[2]. Underlying research data will consist of selected parts of the general datasets generated, and for which the decision of making that part public has been made.

Other datasets will be related to any public report or be useful for the research community. They will be selected parts of the general datasets generated or full datasets (e.g. up to 2 years of key operating data) and be published as soon as they become available.

1.4 ROLES AND RESPONSIBILITIES

Each OPERA 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.

The Project Coordinator has a particular responsibility to ensure that data shared through the OPERA website are easily available, but also that backups are performed and that proprietary data are secured.

OCEANTEC, as WP1 leader, will ensure dataset integrity and compatibility for its use during the project lifetime by different partners.

Validation and registration of datasets and metadata is 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.

Quality control of these data is the responsibility of the relevant WP leader, supported by the Project Coordinator.

[1] **"Gold" Open Access:** Authors make a one-off payment to the publisher so that the scientific publication is immediately published in open access mode.

[2] **"Green" Open Access:** Due to the contractual conditions of the publisher, the scientific publication can undergo an embargo period up to six months since publication date before the author can deposit the published article or the final peer-reviewed manuscript in open access mode.

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 publicly available data. WP1 will provide naming and version conventions.

Last but not least, all partners must consult the concerned partner(s) before publishing data in the open domain that can be associated to an exploitable result.

2. DATA COLLECTION, STORAGE AND BACK-UP

The OPERA project will generate data resulting from instrumentation recordings during the lab testing and open-sea testing. In addition to the raw, uncorrected sensor data, converted and corrected data, as well as several other forms of derived data will be produced.

Instrumentation, data acquisition and logging systems are thoroughly described in D1.1 [2] . A database management system will be used in the project to create, read, update and delete data from a database. The software platform being used is MySQL 5.7.9. A SCADA system will allows partners to access monitoring information locally and remotely.

The following sections describe the different datasets that will be produced in the course of the project.

2.1 ENVIRONMENTAL MONITORING DATA

Environmental monitoring data will be collected at two locations, namely the Mutriku shoreline plant and the open sea test site BiMEP. These numeric datasets will be directly obtained through observations, and derived using statistical parameters and models.

In general, environmental monitoring datasets will be useful for further research activities beyond the scope of OPERA objectives. Metocean observations are common practice for different uses. Dataset could be integrated and reused, particularly for the characterisation of wave resource and the estimation of device performance. They will be also valuable for technology developers who plan to test their devices at either Mutriku or BiMEP.

Although the raw datasets are useful by themselves, it is the objective of the OPERA project to use the data as a basis for at least one scientific publication.

A short description of the environmental monitoring datasets is given next.

TABLE 2.1: WAVE RESOURCE AT MUTRIKU

Reference/Name	• DS_Wave_Mutriku
Description	• Wave resource data 200 m off the shoreline plant. • Main data are the pressure fluctuations over time.
Source	• RBR & Isurki Pressure Gauges
Type	• Observational and derived
Format	• MySQL database for real-time data • TXT for instrument recordings and derived data
Software	• Scilab program to transform pressure into wave height and period. • Spectral analysis software.
Estimated size	• 2 GB (6 months @ 2 Hz sampling frequency)
Storage	• Internal USB memory stick, on-site database server, and real-time replication onto cloud-hosted database server
Back-up	• Daily back-ups on both local and cloud-hosted servers. 15-day retention period for incremental backups in the latter.

TABLE 2.2: WAVE RESOURCE AT BIMEP

Reference/Name	• DS_Wave_BiMEP
Description	• Wave resource at 300 m up-wave of the WEC. • Datasets mainly consist of wave parameters such as wave H_s, T_p, direction and spreading.
Source	• TRIAXYS surface following buoy
Type	• Observational and derived
Format	• MySQL database for real-time data • TXT for instrument recordings • MS Excel for derived and filtered data
Software	• Spectral analysis software
Estimated size	• 150 MB of statistical data (20 min x 2 years) • 1 GB of real-time data (20 min x 2 Hz sampling frequency when real-time communications activated) • 8 GB (2 years @ 2 Hz sampling frequency)
Storage	• Internal USB memory stick, on-site database server and real-time replication onto cloud-hosted database server
Back-up	• Daily back-ups on both local and cloud-hosted servers. • 15-day retention period for incremental backups in the latter.

2.2 MOORING PERFORMANCE DATA

Experimental data will be collected at the DMAC facility in UNEXE [4] . Besides, field tests will be conducted at the open sea test site at BiMEP. Datasets consists of mooring performance data will be both experimental and observational, raw and derived (using statistical parameters and models).

The mooring performance dataset will be useful to inform technology compliance, survivability and reliability as well as economical improvements. They will be also valuable for the certification processes of other technology developers. These data will be the basis for at least one scientific publication.

A short description of the mooring performance datasets is given below.

TABLE 2.3: TETHER LOADS AT DMAC

Reference/Name	• DS_Tethers_Lab
Description	• Characterisation of design load behaviour, fatigue and durability of several elastomeric tether specimens.
Source	• DMAC facility
Type	• Experimental
Format	• CSV (processed data)
Software	• Labview, Optitrack Motive and Matlab
Estimated size	• 8.7 GB (50 Hz sampling frequency)
Storage	• Network storage
Back-up	• Network drive is backed up daily with two-disk fault tolerance (i.e. backups are safe even if two disks fail). Backups are stored in a different building and protected by a dedicated UPS.

TABLE 2.4: MOORING LOADS AT BIMEP

Reference/Name	• DS_Mooring_BiMEP
Description	• Extreme loads and motion response to different sea states will be monitored. • The loading data will be combined with the environmental monitoring dataset to derive the final mooring performance dataset. • Comparison between the polyester lines and the elastomeric mooring tethers.
Source	• MARMOK-A-5 prototype. • A mooring condition monitoring has been implemented for the project consisting of 4 load shackles deployed in two mooring nodes of the prototype.
Type	• Observational and derived
Format	• MySQL database for real-time data • TXT for raw instrument recordings • MS Excel for derived and filtered data
Software	• Statistical and spectral analysis software
Estimated size	• ≤ 400 GB (2.5 years recording x 16 measurements @ 20 Hz)
Storage	• On-site database server and real-time replication onto cloud-hosted database server
Back-up	• Daily back-ups on both local and cloud-hosted servers. • 15-day retention period for incremental backups in the latter.

2.3 BIRADIAL TURBINE PERFORMANCE DATA

Experimental data will be collected at existing IST Turbomachinery Laboratory (Dry Lab) for tests in varying unidirectional flow. Also, field tests will be conducted both at Mutriku shoreline plant and the BiMEP open sea test site. Bi-radial turbine performance data will be both experimental and observational, raw and derived (using statistical parameters and models).

The bi-radial turbine performance dataset will be useful to assess turbine efficiency and reliability. The loading data will be combined with the environmental monitoring dataset to derive the final bi-radial turbine performance dataset.

This dataset will be the basis for at least one scientific publication.

A short description of the bi-radial turbine performance datasets is given below.

TABLE 2.5: BI-RADIAL TURBINE PERFORMANCE AT DRY LAB FACILITY

Reference/Name	• DS_Biradial_Turbine_Lab
Description	• Assess turbine performance though unidirectional steady-state and alternating flow.
Source	• IST Turbomachinery Laboratory. • Sensor data acquired at a frequency of 1kHz for turbine pressure head, plenum temperature and humidity, turbine rotational speed, turbine flow rate and the instantaneous position of the flow control valve. • The voltage and the current of the three AC phases at the input and output of the power electronics were acquired at a frequency of 62.5kHz.
Type	• Experimental
Format	• Matlab “mat” files and comma separated value “csv” text files.
Software	• Matlab (Experimental data acquisition) • A special purpose parallelized C++ software (Data filtering), • A software package written in the Julia language (Computation of the instantaneous and time-averaged turbine shaft power, electrical power and available pneumatic power)
Estimated size	• 320 GB
Storage	• Local PC storage
Back-up	• Static data stored at three computers

TABLE 2.6: BI-RADIAL TURBINE PERFORMANCE AT MUTRIKU

Reference/Name	• DS_Biradial_Turbine_Mutriku
Description	• Assess turbine performance and collect extensive data on drivers of components fatigue such as high rpm and accelerations; electrical, temperature and pressure load cycles; humidity in the cabinet (which exacerbates electrical stress damage); rate of salt accumulation and corrosion.
Source	• Mutriku Wave Power Plant. • Bi-radial turbine-generator set and chamber #9 have been instrumented for the project.
Type	• Observational and derived
Format	• MySQL database for real-time data • MS Excel for derived and filtered data
Software	• Statistical and spectral analysis software
Estimated size	• ≤ 50 GB (6-month recording x 150 measurements @ 4 Hz)
Storage	• On-site database server and real-time replication onto cloud-hosted database server
Back-up	• Daily back-ups on both local and cloud-hosted servers. • 15-day retention period for incremental backups in the latter.

TABLE 2.7: BI-RADIAL TURBINE PERFORMANCE AT BIMEP

Reference/Name	• DS_Biradial_Turbine_BiMEP
Description	• Internal water level, chamber pressure/temperature/humidity, rotation speed and torque to assess turbine efficiency in response to different sea states to compare turbine performance drivers of components fatigue
Source	• MARMOK-A-5 prototype. • Bi-radial turbine-generator set and hull structure have been instrumented for the project.
Type	• Observational and derived
Format	• MySQL database for real-time dataMS Excel for derived and filtered data
Software	• Statistical and spectral analysis software
Estimated size	• ≤ 100 GB (12-month recording x 150 measurements at 4 Hz)
Storage	• On-site database server and real-time replication onto cloud-hosted database server
Back-up	• Daily back-ups on both local and cloud-hosted servers. • 15-day retention period for incremental backups in the latter.

2.4 POWER OUTPUT DATA

Experimental data will be collected at electrical test rigs of UCC [5] and TECNALIA [6] . Besides, field tests data will be collected at Mutriku shoreline plant and at the BiMEP open sea test site. Numerical models will be also used to extend the dataset beyond sea-trials data. In the latter, specialist software may be needed for further processing the data. Selected parts of the generated datasets generated will be made public. Power output data will be both experimental and observational, raw and derived such as mean, standard deviation, minimum and maximum values.

Power output data will be useful to identify sources of uncertainty in power performance prediction and for the certification processes of other technology developers.

This dataset will be the basis for at least one scientific publication. A short description of the power output datasets is given below.

TABLE 2.8: POWER OUTPUT AT ELECTRICAL TEST RIG

Reference/Name	• DS_Power_Output_Lab
Description	• Generator speed, voltage, frequency and electric power.
Source	• Electrical test rigs of UCC and TECNALIA
Type	• Experimental and Simulation
Format	• MS Excel
Software	• MATLAB numerical model of the Mutriku Wave Power Plant
Estimated size	• 20 GB (7 CLs x approx. 300 MB)
Storage	• Network storage
Back-up	• Daily back-ups

TABLE 2.9: POWER OUTPUT AT MUTRIKU

Reference/Name	• DS_Power_Output_Mutriku
Description	• Generator speed, voltage, frequency and electric power, including phase voltages & currents.
Source	• Mutriku Wave Power Plant. • Bi-radial turbine-generator set and chamber #9 have been instrumented for the project.
Type	• Observational and derived
Format	• MySQL database for real-time data • MS Excel for derived and filtered data
Software	• Statistical and spectral analysis software
Estimated size	• ≤ 50 GB (6-month recording x 150 measurements @ 4 Hz)
Storage	• On-site database server and real-time replication onto cloud-hosted database server
Back-up	• Daily back-ups on both local and cloud-hosted servers. • 15-day retention period for incremental backups in the latter.

TABLE 2.10: POWER OUTPUT AT BIMEP

Reference/Name	• DS_Power_Output_BiMEP
Description	• Generator speed, voltage, frequency and electric power, including phase voltages & currents.
Source	• MARMOK-A-5 prototype. Bi-radial turbine-generator set and hull structure have been instrumented for the project.
Type	• Observational and derived
Format	• MySQL database for real-time data • MS Excel for derived and filtered data
Software	• Statistical and spectral analysis software
Estimated size	• ≤ 100 GB (12-month recording x 150 measurements at 4 Hz)
Storage	• On-site database server and real-time replication onto cloud-hosted database server
Back-up	• Daily back-ups on both local and cloud-hosted servers. • 15-day retention period for incremental backups in the latter.

2.5 POWER QUALITY DATA

Experimental data will be collected at electrical test rig of UCC [10]. Also, field tests data will be collected at the Mutriku shoreline plant. Simulated models may be used to assess the power quality for other operating conditions, such as varying control algorithms, resource conditions, grid strengths, and control using a dry-lab to create a wider profile for the WEC. Power quality data will be both experimental and observational, raw and derived (using statistical parameters and models). Selected parts of the experimental datasets generated will be made public.

Power quality data will be useful to identify sources of uncertainty in assessing the impact of the wave energy converter on the performance of the grid. They will be also valuable for the certification processes of other technology developers.

This dataset will be the basis for at least one scientific publication.

A short description of the power quality datasets is given next.

TABLE 2.11: POWER QUALITY AT ELECTRICAL TEST RIG

Reference/Name	• DS_Power_Quality_Lab
Description	• Current, voltage, power quality characteristic parameters (such as voltage fluctuations, harmonics, inter-harmonics, active/reactive power, and flicker).
Source	• Electrical test rig at UCC
Type	• Experimental and Simulation
Format	• MS Excel
Software	• MATLAB Simulink numerical model of the Mutriku Wave Power Plant
Estimated size	• Maximum 1.2 GB per 10-minute test (at 20 kHz sampling frequency). • 4 signals at 20 kHz for 10 minutes per test
Storage	• Network storage
Back-up	• Daily back-ups

TABLE 2.12: POWER QUALITY AT MUTRIKU

Reference/Name	• DS_Power_Quality_Mutriku
Description	• Data will be collected from both a single turbine and the plant as a whole, obtaining valuable conclusions about how aggregation of multiple turbines affects the power quality.
Source	• Mutriku Wave Power Plant.
Type	• Observational and derived
Format	• MySQL database for real-time data • MS Excel for derived and filtered data
Software	• LabView • Statistical and spectral analysis software
Estimated size	• > 200 GB (12-month recording x 12 measurements @ 20 kHz). • Given the large data storage requirements, the measurements will be triggered, and not carried out continuously. After sufficient power quality analysis has been carried out at 20 kHz, the sampling rate will then be reduced (to approximately 12 kHz, and 10 kHz).
Storage	• On-site database server and real-time replication onto cloud-hosted database server
Back-up	• Daily back-ups on both local and cloud-hosted servers. • 15-day retention period for incremental backups in the latter.

2.6 OFFSHORE OPERATIONS DATA

Field tests will be conducted at the BiMEP open sea test site. The offshore operations data will be combined with the environmental monitoring dataset to derive the final dataset. Collected datasets will be made public. Offshore operations data will be observational and derived.

Offshore operations data will be useful to reduce the uncertainty on the determination of risk and cost of offshore operations, and to optimise these activities. The offshore logistics experience can be extrapolated to different scenarios of larger deployment with a view to more accurately assess the economies of scale and identify logistics bottlenecks when deployed in large arrays.

Although the raw datasets are useful by themselves, it is the objective of the OPERA project to use the dataset as a basis for at least one scientific publication.

A short description of the offshore operations datasets is given below.

TABLE 2.13: OFFSHORE OPERATIONS

Reference/Name	• DS_Offshore_Operations
Description	• Failures, type of maintenance, offshore resources (such as vessels, equipment, personnel, parts and consumables), health & safety, and activity log.
Source	• Unlike the previous datasets, these are not based on process instrumentation and therefore will not be stored in the WP1 database.
Type	• Observational
Format	• MS Excel
Software	• n/a
Estimated size	• 10 MB
Storage	• Network storage
Back-up	• Daily back-ups on a separate server

3. DATA STANDARDS AND METADATA

The following standards should be used for data documentation:

- Ocean Data Standards Project [7] : it contains an extensive number of references on Oceanographic Data Management and Exchange Standards. It includes references on Metadata, Date and Time, Lat/Lon/Alt, Country names, Platform instances, Platform types, Science Words, Instruments, Units, Projects, Institutions, Parameters, Quality Assurance and Quality Control.
- ISO 19156:2011 [8] : it defines a conceptual schema for observations, and for features involved in sampling when making observations. These provide models for the exchange of information describing observation acts and their results, both within and between different scientific and technical communities.
- IEC TS 62600-101 [9] : technical specification for wave energy resource assessment and characterisation.
- DNVGL-OS-E301 [10] : it contains criteria, technical requirements and guidelines on design and construction of position mooring systems. The objective of this standard is to give a uniform level of safety for mooring systems, consisting of chain, steel wire ropes and fibre rope.
- IEC TS 62600-10 [11] : technical specification for assessment of mooring system for Marine Energy Converters (MECs).
- IEC TS 62600-100 [12] technical specification on power performance assessment of electricity producing wave energy converters
- IEC TS 62600-102 [13] technical specification on wave energy converter power performance assessment at a second location using measured assessment data
- IEC TS 62600-30 [14] technical specification on electrical power quality requirements for wave, tidal and other water current energy converters
- IEC 61000-4-7:2002 [15] further instructions on processing harmonic current components are given in for power supply systems and equipment connected thereto.
- FRACAS [16] Failure the Reporting, Analysis and Corrective Action System
- ISO 14224:2006 [17] collection and exchange of reliability and maintenance data for equipment.

Metadata records will accompany the data files in order to describe, contextualise and facilitate external users to understand and reuse the data.

OPERA will adopt the DataCite Metadata Schema [18] , a domain agnostic metadata schema, as the basis for harvesting and importing metadata about datasets from data archives. The core mission of DataCite is to build and maintain a sustainable framework that makes it possible to cite data through the use of persistent identifiers.

The following metadata should be created to identify datasets:

- Identifier: A unique string that identifies the dataset
- Author/Creator: The main researchers involved in producing the data in priority order
- Title: A name or title by which a data is known
- Publisher: The name of the entity that holds, archives, publishes prints, distributes, releases, issues, or produces the data.
- Publication Year: The year when the data was or will be made publicly available
- Subject: Subject, keyword, classification code, or key phrase describing the resource.
- Contributor: Name of the funding entity (i.e. "European Union" & "Horizon 2020")
- Size: Unstructured size information about the dataset (in GBs)
- Format: Technical format of the dataset (e.g. cvs, txt, xml, ...)
- Version: The version number of the dataset
- Access rights: Provide a rights management statement for the dataset. Include embargo information if applicable
- Geo-location: Spatial region or named place where the data was gathered

4. DATA SHARING AND REUSE

During the life cycle of the OPERA project datasets will be stored and systematically organised in a database tailored to comply with the requirements of WP1 (for more details on the database architecture, please see D1.1 Process instrumentation definition [2]). An online data query tool was operational in Month 12, and available for open dissemination by Month 18. The database schema and the queryable fields, will be also publicly available to the database users as a way to better understand the database itself.

In addition to the project database, relevant datasets will be also stored in ZENODO [19] [10] , which is the open access repository of the Open Access Infrastructure for Research in Europe, OpenAIRE [20]

All collected datasets will be disseminated without an embargo period unless linked to a green open access publication. Data objects will be deposited in ZENODO under:

- Open access to data files and metadata and data files provided over standard protocols such as HTTP and OAI-PMH.
- Use and reuse of data permitted.
- Privacy of its users protected.

Data access policy is summarised in the following table.

TABLE 4.1: DATA ACCESS POLICY

Dataset	Data access policy
DS_Wave_Mutriku	• Unrestricted since no confidentiality or IPR issues are expected regarding the environmental monitoring datasets • Licence: CC-BY
DS_Wave_BiMEP	
DS_Tethers_Lab	• Restricted to WP2 participants, in order to protect the commercial and industrial prospects of exploitable results (KER1 and KER3). • Samples of aggregated data (e.g. load averages or extreme load ranges) will be shared in the open domain for the most relevant sea states. • Licence: CC-BY-ND
DS_Mooring_BiMEP	
DS_Biradial_Turbine_Lab	• Restricted to WP3 participants, in order to protect the commercial and industrial prospects of exploitable results (KER1 and KER2). • Samples of aggregated data (e.g. chamber pressure, air flow, mechanical power) will be shared in the open domain. • Licence: CC-BY-ND
DS_Biradial_Turbine_Mutriku	
DS_Biradial_Turbine_BiMEP	

DS_Power_Output_Lab DS_Power_Output_Mutriku DS_Power_Output_BiMEP	• Restricted to WP4 and WP5 participants, in order to protect the commercial and industrial prospects of exploitable results (KER1, KER4 and KER6). • Samples of aggregated data (e.g. electric power for the different control laws) will be shared in the open domain. • Licence: CC-BY-ND
DS_Power_Quality_Lab DS_Power_Quality_Mutriku	• Restricted to WP5 participants, in order to protect the commercial and industrial prospects of exploitable results (KER1, KER4 and KER6). • Samples of aggregated data (e.g. active, reactive power and power factor) will be shared in the open domain. • Licence: CC-BY-ND
DS_Offshore_Operations	• The aggregated dataset (e.g. operation time, forecast vs recorded wave conditions) will just be shared in the open domain in order to protect the commercial and industrial prospects of exploitable results (KER1 and KER8). • Licence: CC-NC-BY-ND

5. DATA ARCHIVING AND PRESERVATION

The OPERA project database will be designed to remain operational for 5 years after project end. By the end of the project, the final dataset will be transferred to the ZENODO repository, which ensures sustainable archiving of the final research data.

Items deposited in ZENODO will be retained for the lifetime of the repository, which is currently the lifetime of the host laboratory CERN and has an experimental programme defined for the at least next 20 years. Data files and metadata are backed up on a nightly basis, as well as replicated in multiple copies in the online system. All data files are stored along with a MD5 checksum of the file content. Regular checks of files against their checksums are made.

6. REFERENCES

[1] D8.13 Plan for Exploitation (2017). OPERA Project.

[2] D1.1 Process instrumentation definition (2016). OPERA Project.

[3] UK Data Archive. Research Data Lifecycle. Available at: http://data-archive.ac.uk/create-manage/life-cycle

[4] DMAC: Dynamic Marine Component Test Facility, University of Exeter. Available at: http://emps.exeter.ac.uk/renewable-energy/research/

[5] UCC Beaufort Rotating test rig. Available at: http://www.fp7-marinet.eu/UCC_beaufort-rotating-test-rig.html

[6] TECNALIA electrical PTO Lab. Available at: http://www.tecnalia.com/en/energy-environment/news/electrical-pto-lab.htm

[7] Ocean Data Standards Project. Available at: http://www.oceandatastandards.org/

[8] ISO 19156:2011 Geographic information -- Observations and measurements (2011). ISO.

[9] IEC TS 62600-101:2015 Marine energy - Wave, tidal and other water current converters - Part 101: Wave energy resource assessment and characterization (2015). IEC.

[10] DNVGL-OS-E301 Position mooring (2015). DNVGL.

[11] IEC TS 62600-10:2015 Marine energy - Wave, tidal and other water current converters - Part 10: Assessment of mooring system for marine energy converters (MECs) (2015).IEC.

[12] IEC TS 62600-100:2012 Marine energy - Wave, tidal and other water current converters - Part 100: Electricity producing wave energy converters - Power performance assessment (2012). IEC.

[13] IEC TS 62600-102 Wave energy converter power performance assessment at a second location using measured assessment data (2016, draft). IEC.

[14] IEC TS 62600-30 Marine Energy - Wave, tidal and other water current converters - Part 30: Electrical power quality requirements for wave, tidal and other water current energy converters (2016, draft). IEC.

[15] IEC 61000-4-7:2002 Electromagnetic compatibility (EMC) – Part 4-7: Testing and measurement techniques – General guide on harmonics and interharmonics measurements and instrumentation (ed.2.0, 2002). IEC

[16] Failure Reporting, Analysis and Corrective Action System (FRACAS) Application Guidelines, Product code Fracas. (1999). Reliability and maintenance data for equipment.

[17] ISO 14224:2006, Petroleum, petrochemical and natural gas industries -- Collection and exchange of reliability and maintenance data for equipment. (2010).

[18] DataCite. Available at: https://www.datacite.org/

[19] ZENODO. Available at: http://www.zenodo.org/

[20] OpenAIRE H2020 Project. Available at: https://www.openaire.eu/

[21] Guidelines on Data Management in Horizon 2020, European Commission, Research & Innovation, 2013.

[22] D8.3 Plan for dissemination and communication – update (2017). OPERA Project.

[23] ORCID. Available at: http://orcid.org/

[24] RECOLECTA by FECYT. Available at: http://recolecta.fecyt.es/portada?language=es

ANNEX I: KEY PRINCIPLES FOR OPEN ACCESS TO RESEARCH DATA

These principles can be applied to any project that produces, collects or processes research data. As indicated in Guidelines on Data Management in H2020 [21] , scientific research data should be easily:

1. Discoverable

The data and associated software produced and/or used in the project should be discoverable (and readily located), identifiable by means of a standard identification mechanism (e.g. Digital Object Identifier)

2. Accessible

Information about the modalities, scope and licenses (e.g. licencing framework for research and education, embargo periods, commercial exploitation, etc.) in which the data and associated software produced and/or used in the project is accessible should be provided.

3. Assessable and intelligible

The data and associated software produced and/or used in the project should be assessable for and intelligible to third parties in contexts such as scientific scrutiny and peer review (e.g. the minimal datasets are handled together with scientific papers for the purpose of peer review, data are provided in a way that judgments can be made about their reliability and the competence of those who created them).

4. Useable beyond the original purpose for which it was collected

The data and associated software produced and/or used in the project should be useable by third parties even long time after the collection of the data (e.g. data are safely stored in certified repositories for long term preservation and curation; they are stored together with the minimum software, metadata and documentation to make it useful; the data are useful for the wider public needs and usable for the likely purposes of non-specialists).

5. Interoperable to specific quality standards

The data and associated software produced and/or used in the project should be interoperable allowing data exchange between researchers, institutions, organisations, countries, etc. (e.g. adhering to standards for data annotation, data exchange, compliant with available software applications, and allowing re-combinations with different datasets from different origins).

ANNEX II: SCIENTIFIC PUBLICATIONS

Project Partners are responsible for the publication of relevant results to scientific community by Scientific Publications. According to OPERA DoA, at least 8 indexed manuscripts will be produced. The list of scientific publications is available in Deliverable D8.3 Plan for dissemination and communication [22] .

The data (including associated bibliographic metadata) needed to validate the results presented in scientific publications will be deposited in a research data repository. This requirement is based on the fact that the concept of 'publication' has rapidly evolved over the past years and in the context of the digital era. Therefore, the notion of 'publication' increasingly includes the data underpinning the publication and results presented, also referred to as 'underlying' data. This data is needed to validate the results presented in the deposited scientific publication and is therefore seen as a crucial part of the publication and an important ingredient enabling scientific best practice.

Metadata will maximise the discoverability of publications and ensure the acknowledgment of EU funding. Bibliographic data mining is more efficient than mining of full text versions. The inclusion of metadata is necessary for adequate monitoring, production of statistics, and assessment of the impact of H2020. In addition to basic bibliographic information about deposited publications the following metadata information is expected:

- EU funding acknowledgement:
  - Contributor: "European Union (EU)" & "Horizon 2020"
- Peer Reviewed type (e.g. accepted manuscript; published version).
- Embargo Period (if applicable):
  - End date.
  - Access mode.
- Project Information:
  - Grant number: “654444”
  - Name of the action: “Research and Innovation action”
  - Project Acronym: “OPERA”
  - Project Name: “Open Sea Operating Experience to Reduce Wave Energy Costs”
- Publication Date.
- Persistent Identifier:
  - Authors and Contributors. Wherever possible identifiers should be unique, non-proprietary, open and interoperable (e.g. through leveraging existing sustainable initiatives such as ORCID [23] [23] for contributor identifiers and DataCite [18] for data identifiers).
  - Research Outcome
- License. The Commission encourages authors to retain their copyright and grant adequate licences to publishers. Creative Commons offers useful licensing solutions.

OPERA project will support the open access approach to scientific publications (as defined in article 29.2 of the Grant Agreement). Scientific publications covered by an editorial copyright will be made available internally to the partners and shared publicly through references to the copyright owners web sites.

Whenever possible, a scientific publication, as soon as possible and at the latest six months after the publication time, will be deposited in a machine-readable electronic copy of the published version or final peer-reviewed manuscript accepted for publication in a repository for scientific publications. Moreover, the beneficiary should aim at depositing at the same time the research data needed to validate the results presented in the deposited scientific publications.

TECNALIA has just finalised the development of the TECNALIA Publications repository which is an open access repository accessible by RECOLECTA [24] (a platform which gathers all scientific repositories at Spanish national level) and OpenAIRE [20] (a new platform aimed at gathering a H2020 EU funded-projects' scientific publications). The repository is indexed by Google and fulfils international interoperability standards and protocols to gain long-term sustainability.

All scientific publications of the OPERA project will be available through OpenAire repository which allows searching publications per project. The potential delayed access ('embargo periods') required by specific publishers and magazines will be negotiated in a case-by-case basis.

All research data underpinning a publication will be openly accessible as defined in article 29.3 of OPERA Grant Agreement. Similarly, as with the scientific publications, ZENODO [10] , which is the open access repository of OpenAIRE [20] , will be used principally.