Data Management Plan V1

Document Code: MELOA-DMS-PMD-PMP02-E-R

Date of delivery:	15/11/2018
Deliverable identifier:	D1.02
Version of document:	1.0
Dissemination level for document:	PU
Prepared by:	Diego Lozano
Reviewed by:	Nuno Almeida
Approved by:	MELOA Technical Steering Committee

Project start date:	01/12/2017
Project duration:	39 months

Approval Signature on behalf of Technical Steering Commitee		
Name	Function	Signature
Diego Lozano	Project Coordinator	

List of Participants			
Participant No	**Organization Name**	**Acronym**	**Country**
1	DEIMOS SPACE SLU	DMS	ES
2	COLLECTE LOCALISATION SATELLITES SA	CLS	FR
3	INSTITUTO DE ENGENHARIA DE SISTEMAS E COMPUTADORES, TECNOLOGIA E CIÊNCIA	INESC	PT
4	UNIVERSITAT POLITECNICA DE CATALUNYA	UPC	ES
5	SMARTBAY IRELAND	SBI	IE
6	COMPOSITE SOLUTIONS LDA	CMPSOL	PT
7	INSTITUTO HIDROGRÁFICO	IH	PT
8	INSTITUTO DE SISTEMAS E ROBÓTICA	ISR	PT
9	DEIMOS ENGENHARIA SA	DME	PT
10	OCEANSCAN – MARINE SYSTEMS & TECHNOLOGY SA	OSCAN	PT

Revision Records			
Version	**Date**	**Changes**	**Authors**
1.0	15/11/2018	First version of the document	Diego Lozano

www.ec-meloa.eu

1 Table of Contents

LIST OF TABLES

www.ec-meloa.eu

Table 1 Acronyms and Abbreviations

Acronyms and Abbreviations	
EASME	Executive Agency for Small and Medium-sized Enterprises
EU	European Union
GNSS	Global Navigation Satellite System
MELOA	Multi-purpose/Multi-sensor Extra Light Oceanography Apparatus
SST	Sea Surface Temperature
DMP	Data Management Plan
FAIR	Findable, Accessible, Interoperable and Re-use
CSV	Comma Separate Value
OGC	Open Geospatial Consortium
IMU	Inertial Measurement Unit

Table 2 Reference Documents

Ref	Document Title	Date
R-1	https://guidelines.openaire.eu/en/latest/data/index.html	
R-2	https://creativecommons.org/licenses/by/4.0/legalcode	
R-3	https://creativecommons.org/licenses/by-sa/4.0/legalcode	
R-4	https://opendatacommons.org/licenses/by/	
R-5	https://opendatacommons.org/licenses/odbl/	
R-6	https://www.fiware.org/about-us/	

1 INTRODUCTION

1.1 Project Overview

The MELOA project proposes to develop a low-cost, easy-to-handle, wave resilient, multi-purpose, multi-sensor, extra light surface drifter for use in all water environments, ranging from deep-sea to inland waters, including coastal areas, river plumes and surf zones. The device will be developed as an upgrade to the WAVY drifter conceived by the Faculty of Engineering of the University of Porto, which was used to measure the surface circulation forced by wave breaking, including detailed structure of rifts and the littoral drift current.

The philosophy of the WAVY drifter will essentially be respected:

- a small-size sphere with just enough room to accommodate power source, GNSS-receiver, communications modules, antennae, sensors and data processor;

- optimised buoyancy to prevent the drifter trajectory responding to the wind instead of the current, while providing just enough exposure of the antennae to ensure acquisition of the GNSS signal at the required rate and reliable near real-time communications.

Given the low influence of wind upon the drifters' displacements, MELOA will provide a cheap effective way to monitor surface currents and surface dynamic features anywhere in the World Ocean. Through equipping the drifters with thermistors at two different levels, the possibility is open for monitoring "near-skin temperature" and nearsurface vertical temperature gradients, which will be invaluable for calibration/validation of satellite derived SST fields.

General Information	
Project Title	Multi-purpose/Multi-sensor Extra Light Oceanography Apparatus
Starting Date	1st December 2017
Duration in Months	39
Call (part) Identifier	H2020-SC5-2017-OneStageB
Topic	SC5-18-2017 Novel in-situ observation systems
Fixed EC Keywords	Market development, Earth Observation / Services and applications, Technological innovation, In-Situ Instruments / sensors
Free Keywords	Novel measurements; Cost reduction

1.2 Scope

The Data Management Plan (DMP) will detailing what data the project will generate, whether and how it will be exploited or made accessible for verification and re-use, and how it will be curated and preserved. The purpose of the DMP is to support the data management life cycle for all data that will be collected, processed or generated by the project.

1.3 Responsibilities

The table below provides information on who's contributed to this document and to which sections.

Table 3 Document Responsibilities

Name	Institution	Responsibilities
Diego Lozano García	DMS	All sections
Nuno Almeida	DME	Revision
Félix Pedrera García	DMS	Revision
Jorge Silva	IH	Revision
Joaquin del Rio	UPC	Revision

1.4 Document Structure

This document is structured as following:

- Section 1 provides a project overview and then goes on to describes the scope, responsibilities and structure of this deliverable.

- Section 2 will describe the datasets generated in the project.

- Section 3 will analyse each of the aspects of the FAIR data: Findable, Accessible, Interoperable and Re-use

- Section 4 will present the resources allocation

- Section 5 will deal with data security

2 Data Summary

MELOA will develop a family of five versions of a low-cost, light-weight multiparameter drifters. These WAVY drifters will be very easy to carry around and to deploy. In contrast and in spite of that, the data they produce have far-reaching applications, directly providing valuable information that will help to derive answers to diverse scientific, environmental and societal needs and achieving multiple objectives, from complementing observational gaps in ocean observation, to delivering validation datasets to satellite ground-truthing, along with the real possibility of their effective use by the common citizen.

www.ec-meloa.eu

The data generated in the MELOA project will be acquired in the test campaign and demonstrations of the WAVY drifters. A data set in MELOA is the collection of data samples acquired by a WAVY during a campaign. The contents of the data samples depend on the type of WAVY drifter. The common information for all the types is the GNSS (Time, position, velocity and direction) and the battery power. The table below presents the contents of the data samples for each type of WAVY drifter and the approximate size.

Table 4 WAVY dataset contents

WAVY type	Sensors	Data sample contents	Sample size (approx. in CSV)
WAVY basic	GNSS (1Hz) Thermistor (0.17Hz)	Timestamp, Position, velocity, direction, n. satellites, HDOP (76 bytes) 1x temperature (7 bytes) Battery power (7 bytes)	90 bytes
WAVY littoral	GNSS (1Hz) IMU (20Hz)	Timestamp, Position, velocity, direction, n. satellites, HDOP (76 bytes) wave parameters (Wavelength, Amplitude, Period & Speed) + 5 fourier coefficients (120 bytes) battery power (7 bytes)	203 bytes
WAVY ocean	GNSS (1Hz) 2xThermistors (0.17Hz) IMU (20Hz)	Timestamp, Position, velocity, direction, n. satellites, HDOP (76 bytes) wave parameters (Wavelength, Amplitude, Period & Speed) + 5 fourier coefficients (120 bytes) 2x temperatures (14 bytes) battery power (7 bytes)	217 bytes
WAVY ocean plus	GNSS (1Hz) 2xThermistors (0.17Hz) IMU (20Hz)	Timestamp, Position, velocity, direction, n. satellites, HDOP (76 bytes) wave parameters (Wavelength, Amplitude, Period & Speed) + 5 fourier coefficients (120 bytes) 2x temperatures (14 bytes) battery power (7 bytes)	217 bytes
WAVY ocean atmo	GNSS (1Hz) 2xThermistors (0.17Hz) IMU (20Hz) 1xAir pressure gauge (0.17Hz) 2xThermistors (0.17Hz)	Timestamp, Position, velocity, direction, n. satellites, HDOP (76 bytes) wave parameters (Wavelength, Amplitude, Period & Speed) + 5 fourier coefficients (120 bytes) 2x temperatures (14 bytes) 1x air pressure value (7 bytes) 2x air temperatures (7 bytes) battery power (7 bytes)	238 bytes

The format of the data set is equivalent to the L1 Product format, which is defined in the deliverables WAVY L1 Product Specifications (V1 D4.04, V2 D4.12). In the current version, it basically consists of a CSV file containing the data samples acquired by a WAVY and a metadata JSON file specifying the WAVY id, campaign, time period and location. In further iterations during the project, other data formats like O&M JSON and GeoJSON may be supported.

The size of a data set depends on the WAVY type, the sampling rate and the duration of the WAVY activity during the campaign. The sampling rate varies according to the transmission channel used to receive the WAVY data: some minutes with ARGOS Satellite for WAVY ocean, a few seconds with GPRS for WAVY littoral and 1 second with WIFI (sampling rate of the GNSS). For instance, the size of a data set of a WAVY littoral during a day (assuming a sampling rate of 1Hz) will be around 17 MB in CSV format. Note that the raw data from the IMU (recorded at 20Hz) are used to calculate the wave parameters that are the ones stored (with lower rate) in the data sets offered to the user.

The number of datasets obtained in a campaign depends on the duration of the campaign, the type and number of Wavys. The data obtained from a WAVY could be divided into several data sets for different time intervals.

The Field Test campaigns will be defined in the deliverable Field Tests Campaigns Plan (V1 D6.01, V2 D6.02). The lists of MELOA data sets cannot be identified yet and will be described in the next version of the document, after the validation campaigns. The Field Test Campaigns will be divided into groups by open ocean, Argos V4 (Mediterranean sea) and coast (Portuguese, Irish and Spain). Each of these groups will have a dedicated Test Report with two versions, one for each Test period of the project, they will summarise the obtained results and the validity of the data sets for each test campaign:

• Portuguese Coast Field Tests Report (D6.03 and D6.04)

• Irish Coast Field Tests Report (D6.05 and D6.06)

• Spanish Coast Field Tests Report (D6.07 and D6.08)

• Argos V4 Field Tests Report (D6.09 and D6.10)

• Open Ocean Field Tests Report (D6.11 and D6.12)

3 FAIR data

3.1 Making data findable, including provisions for metadata

The MELOA Catalogue solution is based on CKAN, a tool used by national and local governments, research institutions, and other organisations who manage and publish lots of data collections. Once the data sets are published, users can use its faceted search features to browse and find the data set they need, and preview it using maps, graphs and tables. Each data set is associated to a campaign (CKAN group) and an organisation, so that the user can easily browse among the data sets belonging to a campaign or an organisation.

The MELOA catalogue stores the WAVY data sets with metadata following the OGC Observation&Measurement profile and formatted in JSON. Also, there will be a dictionary of the metadata compliant with INSPIRE (ISO 19115). The metadata shall include the following information:

- a title and description for the data acquisition campaign

- the unique ID of the WAVY

- the location and the date of the launch of the campaign

- all the configurations of all the sensors in the WAVY

As introduced above, the format and naming convention of the data sets is defined in the deliverables WAVY L1 Product Specifications (V1 D4.04, V2 D4.12). The filename of a data set includes the version number that makes unique the data set product.

3.2 Making data openly accessible

The data sets are mainly acquired in the Field Test Campaigns during the MELOA project as described in the section 2. The raw data obtained by the WAVY drifters in the campaigns are revised (e.g. cleaning some spurious samples) and in some cases post-processed (e.g. calculating wave parameters) in order to obtain the final data sets that will be offered to the users. The MELOA data sets will be openly accessible after their revision and sometimes after their publication. So far, no access restriction has been identified to the data sets that will be generated in MELOA. In the case that certain data sets will require any access restrictions, they will be clearly identified and explained in the second version of the Data Management Plan.

To have access to the MELOA data sets, a user has to register in the MELOA web portal. The registration will be free and it will allow the users to use the MELOA web applications, in particular the Catalogue and the Geo portal. The Catalogue allows users browsing and searching the data sets collected by the WAVYs for downloading them or previewing them using maps, graphs and tables. It also provides APIs for the access to the data sets by the MELOA Geo portal and other tools such as federated Catalogues.

The Geo portal provides the capability of visualisation of the MELOA data sets in a way that is easy to find and interpret by the general public. The Geo portal retrieves the data sets from the MELOA Catalogue and offers links to the Catalogue for downloading them. The SW user manuals, version 1 and 2, correspond to the deliverables D4.01/D4.02 for the Catalogue and D4.06/D4.07 for the Geo portal. These manuals will be available online in the knowledge base of the Helpdesk in the MELOA web portal and also as a link in the web applications. The MELOA catalogue and Geo portal will be online accesible in the respective URLs:

- http://catalogue.ec-meloa.eu

- http://geoportal.ec-meloa.eu

In order to get closer to relevant user communities, the metadata of the MELOA data sets will be federated with data hubs such as GEOSS and Copernicus. There will be a metadata link between the WAVY data catalogue and the nextGEOSS catalogue. Furthermore, the link to the Copernicus programme will be assured by linking to the CORDA portal and the EuroGOOS, with the provision of data services based on WMS layers (to be provided by the geoportal in V2).

Also, WAVYs data sets will be accesible in the FIWARE catalogue in order to use them by the FIWARE community in the scope of the FIWARE Lab to test integration of the FIWARE SW components with devices such as the WAVYs. FIWARE is an open source community that generates open data and open source code [R-6]. Opening the WAVY data sets to such communities will open new opportunities of data exploitation to the market.

By the time being, we are not planning to deposit the Wavy's data sets in other repositories. However, the possibility will be analysed further taken into account certified repositories from the registry of Research Data repositories (https://www.re3data.org/), in special the ones supported by the openAIR.

3.3 Making data interoperable

For WAVYs data sets, interoperability with other similar platforms will be achieved by using standard implementations from the Open Geospatial Consortium such as the Observation&Measurements profile, Web Map Service, Web Feature Service and Sensor Observation Service. The metadata of the data sets will be compliant with the OGC Observation&Measurements profile. The OGC WMS/WMTS/WFS will be used to export data sets of L1 WAVYs datasets and added-value products to Copernicus data hubs. The API to connect with the FIWARE platform will be supported by the MELOA Data services. These standards and methodologies are useful to federate the data sets with other data hubs and catalogues (FIWARE, NextGEOSS) that are used by other user communities, organisations or institutions.

It will be considered the use of certified repositories that are supported by openAIR, in that case its requirements shall be taken into account [R-1].

The actual data in the MELOA data sets are formatted in CSV files that are easily readable by many standard tools (e.g. Open Office). The CSV format simply allows sharing the MELOA data sets with other users and researchers. Moreover, data from other sources can be translated to CSV files and combine with the WAVY data using commonly used SW applications. Other formats will be evaluated and eventually implemented during the project, such as O&M JSON or GeoJSON.

3.4 Increase data re-use (through clarifying licences)

The MELOA data sets will have an open license that will require the reusers to give attribution to the source of the data (MELOA project). It is still to be decided if the open license will require that the derived data must be shared with the same license (called share-alike). Instead of creating our own license, we may select an existing open license, possible candidates are: Creative Commons (CC-by [R-2], CC-by-sa [R-3]) or Open Data Commons (ODC-by [R-4], ODbL [R-5]). The license will be indicated in the MELOA web portal.

The data sets offered to the users in the MELOA Catalogue are revised to guarantee their quality and validity before they are published with open access. In general, they will be analysed to check that they satisfy the test objectives and are valid. The time to perform the analysis may include the creation of the report for the associated Field test campaign in which the conclusions of the test are agreed.

The MELOA web portal will be kept operational during the lifetime of the MELOA project, although it will be available later as long as the EGI infrastructure keeps the resources. After the project lifetime, the data sets downloaded from the catalogue will remain re-usable in accordance with the terms of MELOA open license.

4 Allocation of resources

The activities for making data FAIR are covered by the tasks in the WP4, in particular by:

- T4.1 Development of catalogue and data storage component: it implements the discovery and access to the MELOA data sets as well as the connectors to FIWARE and the links to other data hubs (Copernicus and NextGEOSS).

- T4.2 Development data processing component (Level 1): this task defines the metadata and formats of the data sets, which are generated by the data processing component implemented in this task.

- T4.3 Development of Level 1 Data visualisation portal: the component developed in this task is provisioning of OGC WMS/WMTS/WFS layers for external applications.

In this way, the cost of making data FAIR in MELOA was already covered by the estimations done for these tasks and it represents a small part of it.

The resources for the long term preservation have not been discussed yet in the project.

5 Data security

The EGI infrastructure provides redundancy of the HW storage with an availability of 99.4%. A backup policy is defined to store all the servers data (including MELOA data sets and Catalogue databases) in a online secure cloud storage (amazon S3). In case of dramatic lost of data in the MELOA web portal, it will be restored from this backup. The backup may include WM images and periodic snapshots to facilitate the recovery procedure.

For the long term, we have not planned yet to store the data sets in a certified repository. They will be available in the MELOA web portal as long as the EGI infrastructure keeps the resources for the project.

www.ec-meloa.eu

END OF DOCUMENT