Offshore Low-trophic Aquaculture in Multi-Use Scenario Realisation

Grant agreement number: 101094065
Topic: HORIZON-MISS-2021-OCEAN-04-01
Starting date: 1 January 2023
Duration: 48 months

D9.2
Title: Data Management Plan 1

Date of delivery: 2023-06-30

Organization name of lead participant for this deliverable: IMR

"Funded by the European Union, grant agreement NO 101094065. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting authority can be held responsible for them."

Dissemination level		
PU	Public, fully open	x
Sen	Sensitive, limited under the conditions of the grant agreement (only for members of the consortium, incl the Commission Services)	

Deliverable number:	D.9.2
Deliverable title:	Data Management Plan 1
Work package:	9
Lead participant:	IMR

Authors		
Name	Organization	E-mail
Rocío Castaño Primo *	IMR	rocio.castano.primo@hi.no
Giulia Dapueto	ETT SPA	giulia.dapueto@grupposcai.it
Antonio Novellino *	ETT SPA	antonio.novellino@grupposcai.it
Carolina Maiorana	IMR	carolina.maiorana@hi.no
Marte Louise Strømme	IMR	marte.stromme@hi.no
Beatrice Maddalena Scotto	ETT SPA	beatrice.scotto@grupposcai.it
Helge Sagen *	IMR	helge.sagen@hi.no

*Corresponding authors

Keywords
Data Management Plan, Open Access, FAIR

Executive Summary
This document constitutes a first version of the Data Management Plan for the project OLAMUR. It includes a description of the types of data generated and used in the project, the data flow from generation to after the project has ended, provisions for FAIR and Open Access, and security and ethics aspects of the data. This is considered a living document and will be updated throughout the project's lifetime.

"Funded by the European Union, grant agreement NO 101094065. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting authority can be held responsible for them."

Document history		
Version	**Date**	**Description**
1.0	2023-06-30	First version of the DMP, delivered to the EC as D9.2 at M6

Table 1: Document history

List of acronyms	
Acronym	**Meaning**
OLAMUR	Offshore Low-trophic Aquaculture in Multi-Use Scenario Realisation
DMP	Data Management Plan
WP	Work Package
MU-LTA	Multi-Use Low-Trophic Aquaculture
OWF	Ocean Wind Farm
ASV	Autonomous Surface Vehicle
UUV	Untethered Underwater Vehicle
UHI	Underwater Hyperspectral Imager
SDG	Sustainable Development Goals
CBD	Convention on Biological Diversity
IPBES	Intergovernmental science-policy Platform on Biodiversity and Ecosystem Services
UNFCCC	United Nations Framework Convention on Climate Change
BBNJ	Marine Biodiversity areas Beyond National Jurisdiction
GDPR	General Data Protection Regulation
DPO	Data Protection Officer

Table 2: List of some of the acronyms used in this document

"Funded by the European Union, grant agreement NO 101094065. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting authority can be held responsible for them."

Table of Contents

"Funded by the European Union, grant agreement NO 101094065. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting authority can be held responsible for them."

1. Introduction

This deliverable presents the first version of the Data Management Plan (DMP) for the OLAMUR project, assigned under deliverable D9.2 to WP9, with the assistance of WP5. This document will be revised and updated during the project lifetime in the form of updates of this deliverable report. Version history can be found on Table 1.

After a brief summary of the OLAMUR project, in this document we describe the types of data that will be used and generated across the WPs, how these data will be managed internally during the project, and their availability after the project lifetime.

The FAIR principles (Wilkins, *et al*., 2016) and the Open Access to research data guidelines from Horizon 2020's Open Research Data Pilot are the guiding frameworks for OLAMUR's data management. The goal is to maximize the impact of research data as its own valid research output and enhance the dissemination of research results and their reuse by facilitating the access and integration. While openness is the key aspect, some data will need to be restricted for a variety of reasons contemplated by the EC that this deliverable documents.

For clarity, "OLAMUR data" is considered data generated either *de novo* in OLAMUR or derived from public, existing, data sources with changes significant enough to grant the consideration of a new dataset. The reuse of pre-existing data is encouraged and will be appropriately acknowledged. If data collected prior to the OLAMUR project that has not been the subject of a publication, published itself or deposited in an official repository, is used in OLAMUR, it will carry the affiliation of OLAMUR and those prior projects. The emergence and publication of pre-existing data in the framework of OLAMUR is an important outcome of the project.

1.1 The OLAMUR project

The objective of OLAMUR (**O**ffshore **L**ow-trophic **A**quaculture in **M**ulti-**U**se scenario **R**ealisation) is to promote commercially viable and sustainable Multi-Use Low Trophic Aquaculture (MU-LTA) in wind farms or fish farms in offshore waters. OLAMUR aims to bring together the existing state-of-the-art practices in MU-LTA related key sectors, to develop and demonstrate a sustainable solution for commercial MU-LTA in both low and high salinity, high eutrophic and high energetic offshore waters. The relevant data, information, products and standards for establishing, operating and evaluating such a solution will be monitored, simulated, stored and customized as an "OLAMUR digital MU-LTA farm service". Through a holistic and interdisciplinary approach, the project aims to demonstrate the possibilities of co-use of marine space and how MU-LTA can contribute to more resilient and sustainable food production with low impacts and emissions.

Services will be developed to support policymakers in making strategic decisions and entering into innovative governance arrangements. This will support producers throughout the entire

"Funded by the European Union, grant agreement NO 101094065. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting authority can be held responsible for them."

chain: planning, implementation, operation, production, impact assessment, commercialization, and capacity building. It will also provide useful information to estimate the potential and perspective of co-localized low-impact aquaculture.

OLAMUR will focus on three geographically and ecologically diverse pilot sites in the Baltic Sea and North Sea to test its efficiency and replicability (Figure 1).

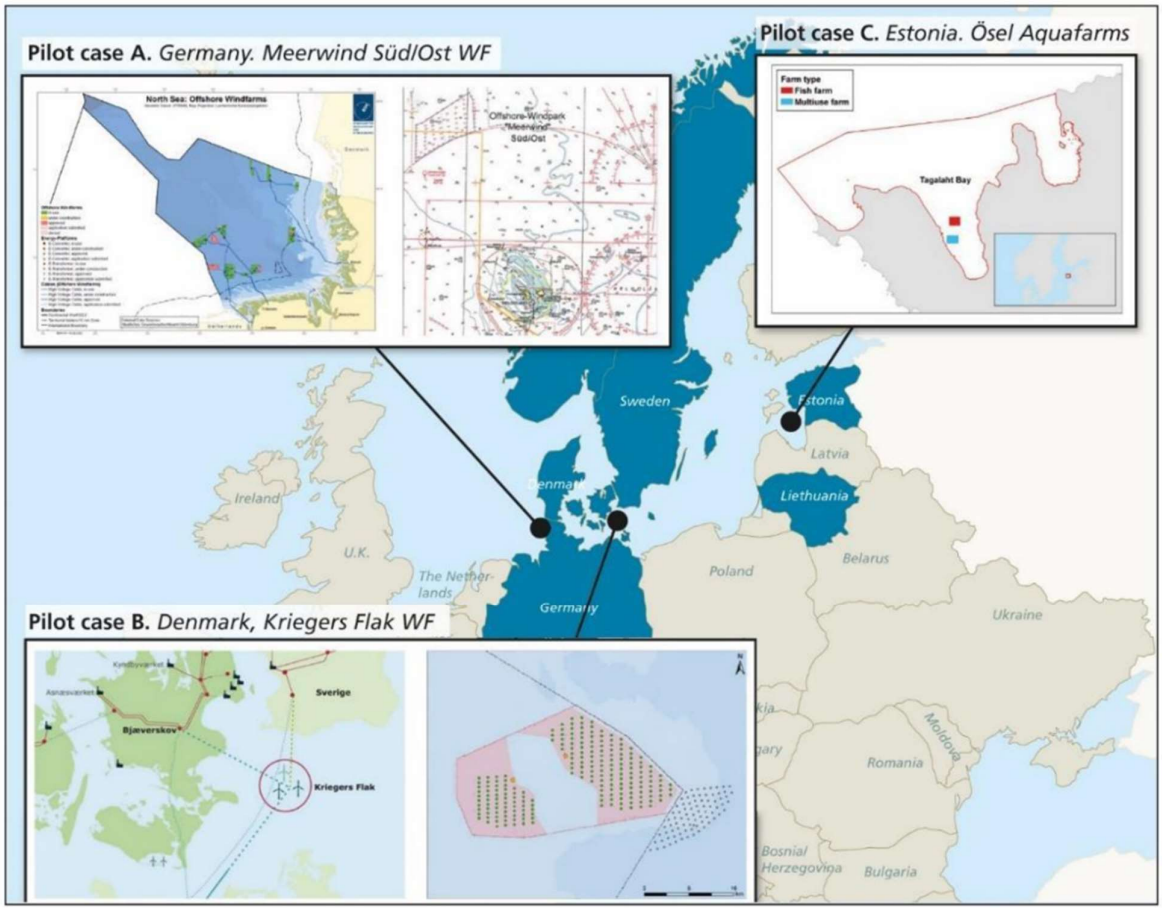

Figure 1: Location of the three OLAMUR pilot sites in the North and Baltic Sea.

Figure 2 shows the OLAMUR research and activities structure and addressed goals: WP2, 3 and 4 are closely related to the technical activities in WP1 whilst WP6 and 7 cover aspects like sustainability and governance; WP5 consolidates the data and makes it available for the project participants and relevant stakeholders. Each WP is organized in tasks and each task is either producing, assessing or developing a data set/product.

The multidisciplinary approach and analysis at diverse levels (planning, implementation, evaluation) implies a wide variety of data types and disciplines: marine environment, aquaculture productivity, societal aspects, technical developments; observations, models, qualitative data.

"Funded by the European Union, grant agreement NO 101094065. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting authority can be held responsible for them."

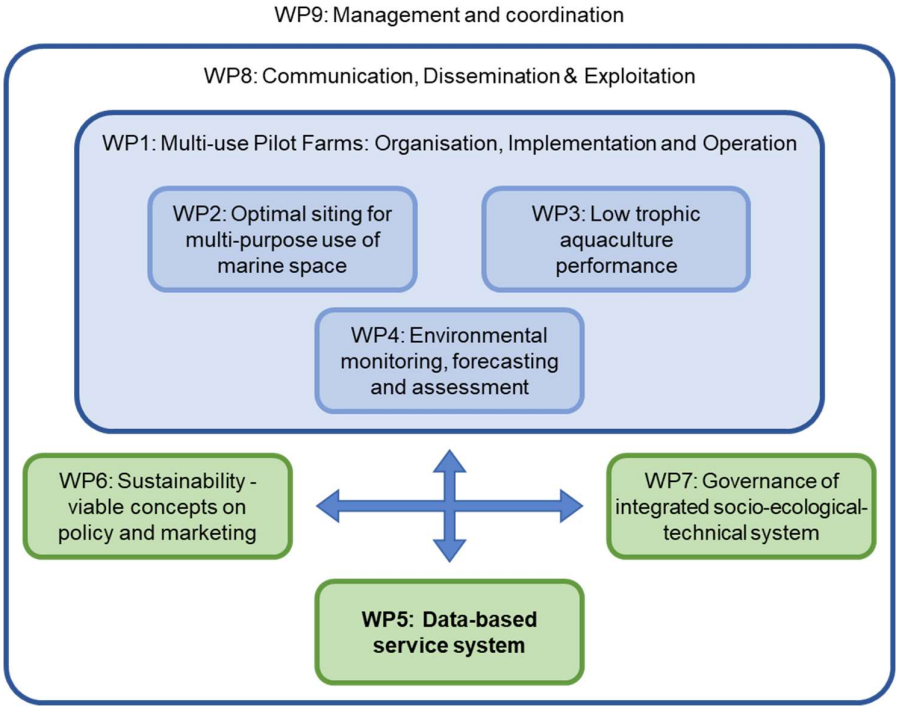

Figure 2: Structure and relationships of OLAMUR's Work Packages

The flow of OLAMUR data within the project is relatively complex; data produced in WPs will be shared across to reach the goals, in some cases in an iterative manner. For example, data from WP1 contributes to the analysis in WP2-4 and the results from WP2-4 will inform new setups in WP1; model output from WP3 and 4 will be used by WP2 to refine the micrositing and regional tools, and WP6 assessments. There will be an intense reuse of others' data within the project, making harmonization efforts crucial from the beginning.

2. Data used and generated in OLAMUR

The OLAMUR project includes many diverse aspects of MU-LTA with many inter-dependencies: logistics and operations of test sites, production optimization, regional assessments, multi-level governance, among others. As a result, the project will generate and use a wide array of different types of data from different fields of research. This chapter provides a description of the types of data that will be generated and used in OLAMUR, with indications of their provenance and their further use within the project.

A substantial amount of pre-existing data used in OLAMUR is made available by marine data aggregators, like Copernicus Marine Service (CMS) and EMODnet, and discipline-specific databases.

WP1 – Multi-use Pilot Farms: organization, implementation and operation

The three pilot sites are central to the OLAMUR project, providing a test setting to experiment and obtain real data on the performance and sustainability of MU-LTA. All

"Funded by the European Union, grant agreement NO 101094065. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting authority can be held responsible for them."

the WPs will use the data gathered at the pilot farms, and in turn the pilot farms will test different iterations based on WP2-4 outputs.

WP1 is centered around the planning, design, construction, operation and maintenance of the pilot farms, while producing data that will be used by other WPs: environmental variables like temperature, salinity, nutrients, light, productivity of mussels and seaweed and their CNP compositions, concentration of harmful metals, and eDNA to account for biodiversity. The design process itself using the OrcaFlex tool will generate engineering data (tension, forces…) that will inform the choice of setup.

WP2 – Optimal siting for multi-purpose use of marine space

The main objective is to develop protocols and tools for determining future optimal co-location of OWF and LTA. At small scale, it will estimate the most suitable setups for seaweed and mussel farms within the OWF (micrositing, T2.2) taking into account physical, economic and regulatory aspects. There will also be a larger, regional assessment of potential areas for MU-LTA in the North and Baltic Seas by extending the existing tool ODSS (Operational Decision Support System, http://www.sea.ee/bbg-odss/Map/MapMain to the North Sea (T2.3).

Due to its geographical focus, WP2 will use and generate data mostly in the form of maps. Kriegers Flak will be the focus of micrositing efforts. Input data are high resolution maps of the physical settings of the wind farm (foundations, cables, zoning, artificial reefs), bathymetry, natural habitats, shipping traffic, historic and current (from WP4) ocean and atmospheric conditions (currents, wind, temperature, salinity), LTA production potential (model output from WP3). It will produce in turn high-resolution maps with options of co-use. These maps will aid WP1 to improve the LTA farm design.

ODSS produces tables that summarize the physical features (sediments, temperature, waves), current human uses (pipelines, ship traffic, fishing effort) and estimated mussel and algal growth and nutrient removal for the area determined by the user. In order to extend the tool to the North Sea, it requires spatial data on seaweed and mussel production and ecosystem services potential (from WP3), environmental data on bathymetry, oceanography, etc. (from existing datasets and WP4) and output of the hydrodynamic model of dispersal OpenDrift (https://opendrift.github.io/). Output from T2.3 provides input to the sustainability analysis in WP6.

The co-development of the protocol for multi-use (T2.1) will not only include the aforementioned physical and biological constrains, but also the stakeholders' perspective in the form of qualitative data gathered from interviews.

WP3 – Low trophic aquaculture performance

WP3 will estimate the potential farming production and the ecosystem services provided by MU-LTA, accounting for the impact of different farming setups and the risk of disease and invasive species. As part of the performance evaluation, the food

"Funded by the European Union, grant agreement NO 101094065. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting authority can be held responsible for them."

safety aspect will be assessed. The impact of artificial reef structures on biodiversity will be investigated in Kriegers Flak.

WP3 uses fine resolution, hydrodynamic-biogeochemical numerical models for each of the test sites: SCHIM in Meerwind, FlexSem-ERGOM in Kriegers Flak and DEB-IBM in the Estonian site. These models require data for constraining, forcing and validation. These data come from public sources, other WPs and other tasks within WP3. In situ observations will be collected for evaluation of contaminants and biodiversity.

T3.1 generates estimates of production of seaweed and mussels in terms of biomass, nutrients and CO_2 uptake and mortality. This output will in turn be used by T3.2 together with CNP composition analysis data from the sites, to quantify ecosystem services in the form of nutrient removal, oxygen generation and water clarity. Data on invasive species will be used to perform a risk assessment. T3.3 will analyze samples from the pilot sites for contaminants and particulate emissions.

WP3 will collect data from the isotopic analysis of otoliths, biological sampling (weight, length, sex, stomach contents, liver and muscle samples) and capture-tag-recapture of Atlantic cod. Together with historic data on cod biodiversity from NDMM and oceanographic, imaging and eDNA data from WP4, they will form the basis to evaluate the impact of artificial reefs (T3.4) on production, metabolic rate and population size of cod.

The output from WP3 will be used by WPs 2 (site selection) and 6 (sustainability assessment).

WP4 – Environmental monitoring, forecasting and assessment

The objective of WP4 is to design an efficient, robotics and model-based monitoring program of MU-LTA sites, carry out data collection accordingly, and develop forecast systems that resolve the impact of OWF and LTA in local environmental conditions.

Robotic and autonomous platforms like drones, ferryboxes, ASV, UUV, moorings, will collect field oceanographic data (temperature, salinity, waves, currents, chlorophyll, light), UHI and stereo-camera images and make them available to the consortium, particularly WP2 and 3, via WP5 (T4.3).

T4.4 to 6 will parametrize the existing operational models SCHISM-WWM (North Sea) and HARMONIE-HBM-WAM (Baltic Sea) to account for the effect of farms. The model output data on local winds, waves and currents, at different configurations, will be compared with the output of the models without the farm parametrization. Coupling with the DEM-IBM model from WP3, WP4 will estimate the effect on future sediment chemistry, oxygen and nutrient uptake, plankton predation, shading, and the ecological carrying capacity of the pilot sites. Environmental data necessary for the models' parametrizations will come from publicly available sources and T4.3, the scenarios for co-location from WP2.

"Funded by the European Union, grant agreement NO 101094065. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting authority can be held responsible for them."

WP5 – Data-based service systems

WP5 oversees the data management in OLAMUR, including making data available within and outside the consortium and ensuring FAIRness and sharing across WPs. Developing and providing tools as such, WP5 is not generating datasets per se but facilitating the exchange, use and further dissemination of data. WP5 uses data and models of other WPs in order to realize software instruments (data service and modules) that will be available to partners, and eventually to other users, to facilitate the multi-use management.

WP6 – Sustainability – viable concepts on policy and marketing

The economic aspects of MU-LTA are addressed by WP6. It will evaluate the economic viability for all actors involved, sustainability aligned with SDG targets, and quantify the benefits beyond financial profits of the three pilot sites.

Using the data generated by WPs 1-5 (farm designs, environmental variables, estimated productivity and ecosystem services…) WP6 will present life cycle sustainability assessments on representative value chains (T6.1) quantifying sustainability measures for carbon, land, water and health footprint (T6.2), Net Present Value of marketed and non-marketed services (T6.3), and via projections (T6.4) assess the contribution of MU-LTA to decarbonizing the European economy via dietary change.

WP7 – Governance of integrated socio-ecological-technical system

WP7 investigates the social and governance side of MU-LTA across the science-policy-industry-community interface (quadruple helix). Adequate governance structures at multiple levels (local to global) and participation of stakeholders are crucial to ensure compliance with regulations after the technical implementations are in place.

Data from WP7 will come primarily from the output of workshops with stakeholders about the relevance of SDGs to their test site, and analysis of texts, interviews and participation in high-level meetings. SINTEF is curating a database of transcripts of CBD, IPBES and UN BBNJ negotiations, which will be used in OLAMUR, and OLAMUR will contribute to.

WP8 / WP9 - Communication, Dissemination and Exploitation / Management and coordination

These two WPs are focused on disseminating the results from the other WPs and the administrative running of the project. They will not generate research data. However, they will deal with personal data (names, affiliations, email addresses) for use in mail lists for internal and external communications, participation in meetings and workshops, feedback forms, etc.

"Funded by the European Union, grant agreement NO 101094065. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting authority can be held responsible for them."

3. Data flow

"Data flow" can be understood as who is using data from whom, or the series of "places" where the project data is to be hosted. The scientific use of data across WPs has been touched in section 2, the WPs' descriptions.

The technical infrastructure for data sharing and distribution within and outside the consortium during the project's lifetime is developed and maintained by WP5 (the OLAMUR backend, Figure 3), including further distribution to European marine data aggregators. The long-term stewardship of data, however, will fall on and be guaranteed by existing, well-established repositories.

Conceptually and in the infrastructure, there will be private areas, accessible only to the consortium, and public areas, where access is granted to the general public without restrictions. The private areas will host data being actively worked on, prior to publication, and data that must remain restricted. The public areas include the public-facing OLAMUR ERDDAP and GeoServer servers, and further distribution of data to European aggregators and long-term archives.

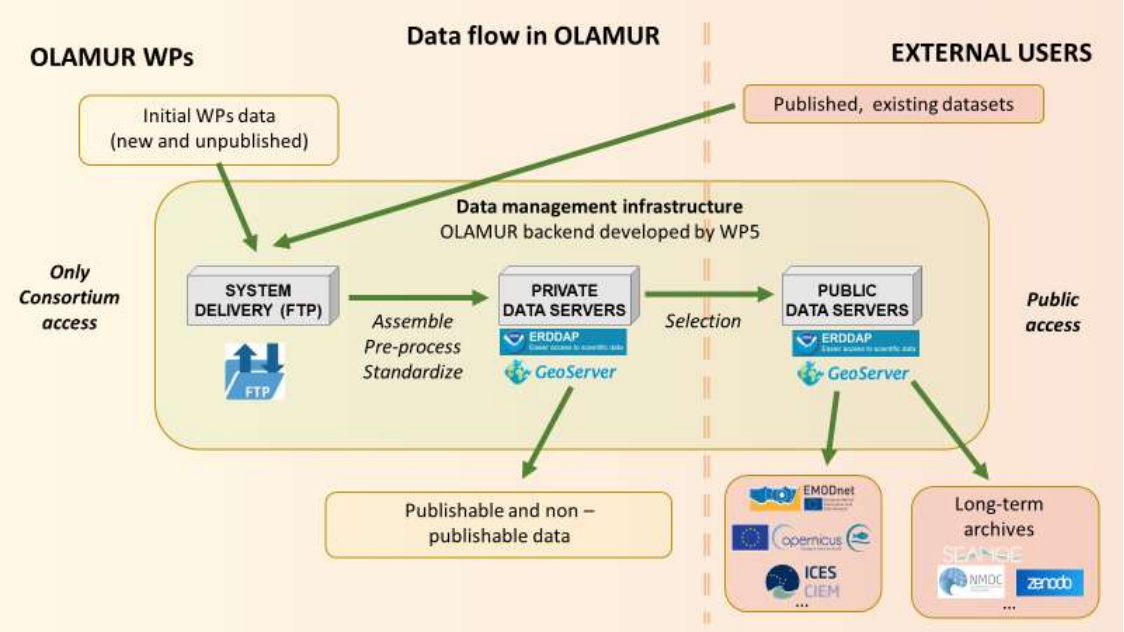

Figure 3: Schematic of OLAMUR data flow

3.1 Data management architecture (OLAMUR backend)

Data produced and used (whether generated by other WPs or other publicly available data like open data from EU meteo-ocean observations and forecasts initiatives) in OLAMUR will be timely assembled and pre-processed with interoperable international standards to the extent possible, to form an OLAMUR data-system and its backend infrastructure. This system will hence not only host the data, but also enforce interoperability from early in the process.

"Funded by the European Union, grant agreement NO 101094065. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting authority can be held responsible for them."

The data management infrastructure is being set up and deployed in terms of resources to host and run data management tools. Moreover, a series of microservices that enable system modularity and scalability, will manage the data harvesting and organization into data catalogues. Specifically, WP5 develops a system delivery where to upload produced and input data via FTP, and a data server where to query and download data from existing databases and from other WPs. In particular, with regard to WP data, once the data is uploaded, it will be transferred to the data server. Regular back-ups of project data and digital information and product results will be performed on external storage media or online solutions.

To facilitate the data harmonization and operate as integrator and data translator for facilitating the data use and interoperability it is important to use common open tools to query for and view data collections and data products. The GOOS Observation Coordination Group (OCG), coordinating the activities of the global ocean observing networks, is working to improve data interoperability between and within the various observing networks.

The OCG is actively promoting the use of ERDDAP (Simons, 2022) as a key tool towards interoperability of global ocean datasets. ERDDAP data server is open-source software written in Java that builds upon the open-source ideals of the OPeNDAP, WCS, SOS and OBIS standards. ERDDAP data server supports several common data file formats (html table, netcdf, csv, txt, json, etc.) and output files are created on-the-fly in any of these formats. ERDDAP implements FGDC Web Accessible Folder (WAF) with FGDC-STD-001-1998 and ISO 19115 WAF with ISO 19115-2/19139.

Another tool that is increasing in popularity and use is GeoServer. GeoServer Java-based server is an open-source software that allows users to view and edit geospatial data. It is built on GeoTools, an open-source Java GIS toolkit. Using open standards set forth by the Open Geospatial Consortium (OGC), GeoServer allows for great flexibility in map creation and data sharing: Web Map Service (WMS) standard allows to create maps in a variety of output formats quickly and easily, thanks also to the integration with OpenLayers, a free map library; Web Feature Service (WFS) and Web Coverage Service (WCS) standards permit the sharing and editing of the data that is used to generate the maps; Web Map Tile Service (WMTS) standard to split your published maps into tiles for ease of use by web mapping and mobile applications. Both ERDDAP and GeoServer are widely adopted and used for marine and atmospheric data management and they both offer easy interoperability with spatial planning and monitoring tools. OLAMUR is adopting data interoperability infrastructure that is based on ERDDAP (to manage tabular and grid data) and GeoServer (to manage vectorial data and map layers). These services allow metadata to be harvested and indexed.

In order to add the data to the ERDDAP and GeoServer instances, both for produced and pre-existing data these fields must be present and standardized as follows:

- Survey area: The survey area can have the following values: "Baltic Sea", "North Sea", "Global Ocean", "Case study A", "Case study B", "Case study C".
- Synthetic name of datum.

"Funded by the European Union, grant agreement NO 101094065. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting authority can be held responsible for them."

- Description: for pre-existing data, the description must be the original title.
- Source: must be either “OLAMUR + WP number” (e.g., “OLAMUR WP5”) or the external source or project (e.g. “Copernicus”).

For data streaming from other ERDDAP instances, the title can’t be changed and will remain the same as the original ERDDAP.

The current OLAMUR backend design considers 2 Virtual Machines (VMs): 2.10GHz - 4 Core, 32GB, running CentOS and 500GB of allocated disk size. The VMs are running Docker containers. Containers are similar to virtual machines (VM), but they have relaxed isolation properties to share the Operating System (OS) amongst the applications. Similar to a VM, a container has its own file system, share of CPU, memory, process space, and more. As they are decoupled from the underlying infrastructure, they are portable across clouds and OS distributions. The most popular solutions are Kubernetes and Docker that have become the de-facto standard and dominant tool with which applications are containerized, managed, scaled and released. As anticipated, OLAMUR adopted Docker that offers packages with service tools like selected data publishing services.

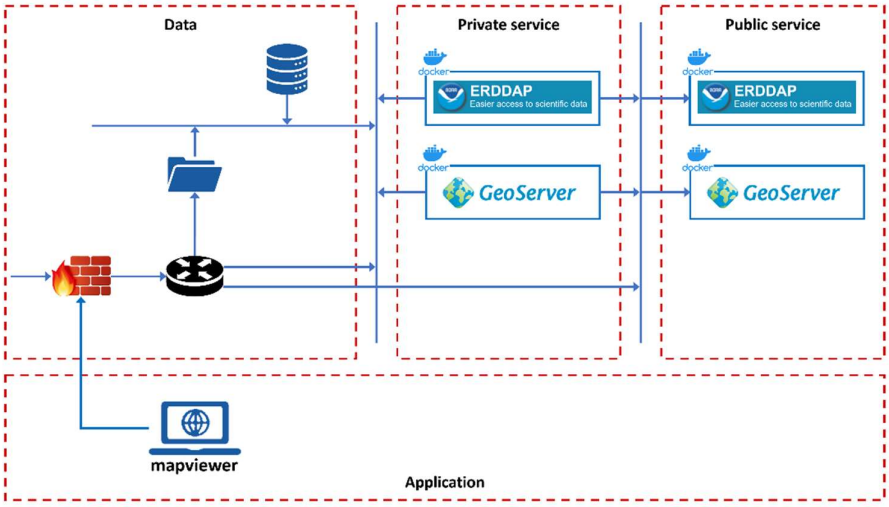

Figure 4: Deployed architecture in the OLAMUR backend.

The management of the access to the private area is implemented by using a selective endpoint access from a given (white) list of IP addresses or host and domain name (fully qualified domain name, FQDN), as shown in Figure 4. Connections and access from edge devices out of the list will be rejected. The idea to design a delivery buffer is mutated from the CMEMS Dissemination Unit (in charge of the dissemination of data products by means of several interfaces and performance monitoring) Delivery buffer. Similarly, the OLAMUR Delivery buffer provides the partners with virtual folders in which new data can be dropped. A preliminary approach suggests the partners to use an internal naming convention for new datasets. This facilitates the successive ingestion step that pushes new data into the OLAMUR database and the data publication tools.

“Funded by the European Union, grant agreement NO 101094065. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting authority can be held responsible for them.”

4. Ensuring FAIRness

The FAIR Guiding Principles for scientific data management and stewardship (Wilkinson *et al*. 2016) are a series of general principles aimed at increasing the reuse of scientific data with a particular emphasis on machine actionability. They cover aspects of Findability, Accessibility, Interoperability and Reusability at the metadata and data level. The principles are as follows:

To be Findable:

F1. (meta)data are assigned a globally unique and persistent identifier

F2. data are described with rich metadata (defined by R1 below)

F3. metadata clearly and explicitly include the identifier of the data it describes

F4. (meta)data are registered or indexed in a searchable resource

To be Accessible:

A1. (meta)data are retrievable by their identifier using a standardized communications protocol

A1.1 the protocol is open, free, and universally implementable

A1.2 the protocol allows for an authentication and authorization procedure, where necessary

A2. metadata are accessible, even when the data are no longer available

To be Interoperable:

I1. (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.

I2. (meta)data use vocabularies that follow FAIR principles

I3. (meta)data include qualified references to other (meta)data

To be Reusable:

R1. meta(data) are richly described with a plurality of accurate and relevant attributes

R1.1. (meta)data are released with a clear and accessible data usage license

R1.2. (meta)data are associated with detailed provenance

R1.3. (meta)data meet domain-relevant community standards

"Funded by the European Union, grant agreement NO 101094065. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting authority can be held responsible for them."

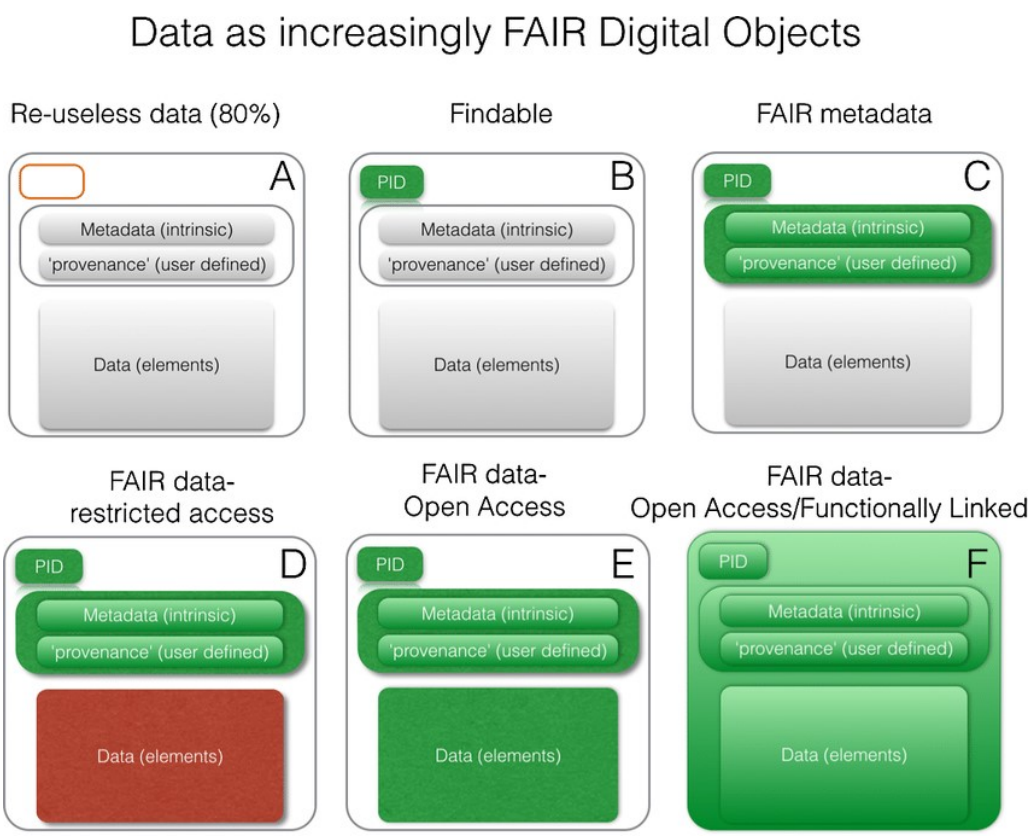

Figure 5: Varying degrees of FAIRness. From Mons et al. (2017)

FAIR data exist in a continuum, with different degrees of FAIRness and room for improvement see Figure 5). Not all fields of research present the same level of FAIR maturity. Fields that generate large amounts of data by "machines", where global collaboration is at the basis of research, or require significant infrastructure, tend to be more FAIR. Genome research (GenBank), remote sensing, weather forecast or climate modelling (CMIP) are examples of these kinds of disciplines. The principle of Interoperability is particularly challenging to achieve. It requires, among other aspects, wide community agreements on standards and a certain level of governance to maintain them.

It is out of the scope of OLAMUR to bring all the disciplines present in the project up to the same level of FAIRness. We will make use of existing, community-agreed systems and approaches, and where they are lacking, the project will strive to maintain a level of intra-project FAIRness. The technical side of the FAIR principles will be in part covered by the long-term archives used to publish the data (minting of identifiers, indexation, metadata standards, communication protocols).

4.1 Making data Findable with provisions for metadata.

The datasets made available by OLAMUR will receive a globally unique, Persistent IDentifier (PID) from the data repository. Digital Object Identifiers (DOI) are the most common PIDs, although the handle system (hdl) is used by some repositories. Other types of PID exist, but they are rarely used for research data. PIDs are globally resolvable, point to a single digital

"Funded by the European Union, grant agreement NO 101094065. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting authority can be held responsible for them."

object and are independent of location. The OLAMUR ERDDAP and Geoserver servers will not mint PID, since they are meant to be temporary, but as data is archived and published, the PID information will be added to the metadata.

The data will be richly described with query metadata to improve findability. To the extent possible, standard metadata fields and keyword vocabularies will be used instead of descriptions and free text. The OLAMUR project will be acknowledged via general keywords, project metadata fields or included in the description, depending on the options the data repository offers. If applicable, version numbers will be provided for provenance. If the data is associated with a publication, the PID of the publication will be included, too. OLAMUR ERDDAP and GeoServer services allow metadata to be harvested and indexed.

Some data formats are self-describing (e.g., NetCDF, GeoPackage) and store both data and metadata in the same file. Where these types of files are used, they will include rich metadata in the file, in addition to the metadata schema of the repository. If possible, the file metadata will include its PID.

4.2 Making data Accesssible

With exceptions that will be justified on a case-by-case basis (see section 5.1), OLAMUR data will be made public without restriction by default, following the H2020 and Horizon Europe guidelines on research data. Data will be deposited in repositories that comply with the FAIR principles. This way the machine accessibility of the data and metadata will be granted by the repositories. If data needs to be pulled out, a strong case must be made and the landing page of the dataset will be substituted by a "tombstone" page that describes the data and the reason for withdrawal. Different versions of the same dataset will link to each other.

Keeping with the spirit of the FAIR principles, OLAMUR will keep the data access threshold as low as possible. Data repositories that do not require registration are preferred. The exceptions are the European data portals like EMODnet and CMS, where login access facilitates tracking of use and data downloads.

4.3 Making data Interoperable

The major principle in data interoperability is that data exchange and reuse and the potential for re-combinations with different datasets from different origins is maximized. The interoperability must be at the machine-level, to allow integration with other data and workflows without the need of a human operator for interpretation and "translation". Appropriate, community-agreed controlled vocabularies are going to be used for metadata and data description. Controlled vocabularies consist of lists of standardized terms, in many cases discipline- or element-specific. Using standardized sets of terms solves the problem of ambiguities associated with data mark-up and also enables records to be interpreted by computers and added value services. Table 3 summarizes some of the controlled vocabularies and standard formats currently established and used in the marine domain. Where multiple options are available, preference will be given to those under European governance, to lower the threshold of further dissemination to European data portals like EMODnet or CMS.

"Funded by the European Union, grant agreement NO 101094065. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting authority can be held responsible for them."

Item	Name, link, other info	Governance
Platform ID	OceanOPS ID / WMO: https://www.ocean-ops.org/ SDN C17: http://vocab.nerc.ac.uk/collection/C17/current/	OCEANOPS WMO SeaDataNet
Platform type	SDN L06: http://vocab.nerc.ac.uk/collection/L06/current/	BODC
Sensor models	SDN L22: http://vocab.nerc.ac.uk/collection/L22/current/	BODC NVS
Institution	EDMO: https://edmo.seadatanet.org/ ROR: https://ror.org/	SeaDataNet
Data mode	NRT/DM/REP	EuroGOOS DATAMEQ
Variable names	SDN P02: http://vocab.nerc.ac.uk/collection/P02/current/ SDN P01: http://vocab.nerc.ac.uk/collection/P01/current/ SDN P07: http://vocab.nerc.ac.uk/collection/P07/current/ Climate and Forecast (CF) standard names: https://cfconventions.org/Data/cf-standard-names/current/build/cf-standard-name-table.html	BODC NVS CF
Unit	SDN P06: https://vocab.nerc.ac.uk/collection/P06/current/	SeaDataNet
Quality Flag Scheme	SDN L20: https://vocab.nerc.ac.uk/collection/L20/current/ WOCE QC flags IOC recommendations: http://www.ioccp.org/images/D4standards/IOC-OceanDataStandards54-3-2013.pdf	SeaDataNet WOCE
Keywords	Global Change Master Directory keywords (GCMD): https://www.earthdata.nasa.gov/learn/find-data/idn/gcmd-keywords	NASA
Time	ISO8601	ISO
Datum	WGS84	ISO
Country	ISO3166	ISO
Licence	Creative Commons: https://creativecommons.org/	CC
INSPIRE	ISO 19115	ISO INSPIRE
PI	ORCID https://orcid.org/	ORCID
Species	World Register of Marine Species (WoRMS): https://www.marinespecies.org/	LifeWatch
Discovery metadata	Attribute Convention for Data Discovery (ACDD): https://wiki.esipfed.org/Attribute_Convention_for_Data_Discovery	ESIP
Atmosphere, ocean data (specially gridded)	NetCDF-CF: https://cfconventions.org/	CF
Ocean mooring data	OceanSites (and derived): http://www.oceansites.org/data/index.html	OceanSITES
Map data	Shape files: https://www.esri.com/content/dam/esrisites/sitecore-archive/Files/Pdfs/library/whitepapers/pdfs/shapefile.pdf	Esri
Biological data	DarwinCore Archive (DwCA): https://dwc.tdwg.org/	TDWG

"Funded by the European Union, grant agreement NO 101094065. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting authority can be held responsible for them."

Table 3: Controlled vocabularies, identifiers and data format standards available to use in OLAMUR

The OLAMUR backend will kick-start the interoperability of data, by harmonizing the data provided by WPs prior to integration into the ERDDAP and GeoServer instances. If adequate vocabularies and formats exist (e.g., oceanographic data), they will be used. If there is a need for a specific format or data requirement from the users, the standards can be adjusted to specific needs enhancing impact.

In terms of file format, data files must not be tied to proprietary software, opting for open formats or formats that can be read and used easily and efficiently by open software.

4.4 Increase data Re-use

Data reuse requires terms of use, and machine-readable data licenses like CC-BY 4.0 (Creative Commons - Attribution) are becoming the *de facto* default usage license for scientific research data. All repositories (to our knowledge) that OLAMUR could potentially use implement CC licenses. CC-BY specifically reflects the academic practice of source citation and allows for a less restrictive reuse, including the use in larger data synthesis. Particular effort will be put in adequately citing datasets in papers and communications, following the FORCE11 Joint Declaration of Data Citation Principles (FORCE 11, 2014).

Unless appropriate reasons can be stated, the usage license should not be more restrictive than CC-BY (e.g. NC (Non-Commercial), and SA (Share-Alike)). While seldomly used, CC0 (public domain) can be used if the dataset creators choose to.

5. Open Access

Open access refers to the practice of providing online access to scientific information that is free of charge to the end-user, and reusable. In the context of research and innovation, 'scientific information' can mean peer-reviewed scientific research articles (published in scholarly journals), or research data (data underlying publications, curated data and/or raw data) (European Commission, 2016). The Open Access procedures to articles is described in the Grant Agreement with the options of “Green” and “Gold” Open Access. Where possible, “Gold” Open Access is preferred, and funds are available in the project for publication fees.

Open Access to OLAMUR research data will be carried out in two ways. During the project lifetime, the public-facing ERDDAP and GeoServer servers (see section 3.1) will make data available without restrictions. After the project ends, publishing in data repositories will ensure future visibility, identification and long-term preservation. Datasets used in scientific publications will be made long-term available by publication time, and adequately cited. Data not used in publications but generated within the OLAMUR project will be archived and published by the end of the project at the latest. It is possible to set up an embargo period for particular cases, like articles in the process of publication or data to be used by a PhD student, but it must be of fixed duration and in no case last beyond the lifetime of the project.

“Funded by the European Union, grant agreement NO 101094065. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting authority can be held responsible for them.”

There are many long-term data repositories, with different target users and levels of service they provide. OLAMUR will not enforce any particular repository, and the researchers are free to choose, as long as it fulfills the FAIR principles and the guidelines from this DMP. To facilitate the choice, we suggest a decision process as follows:

- Their own institutional repositories if they are required by their host institution. E.g., NMDC, BODC, Seanoe.
- Discipline-specific repositories. E.g., Pangaea, many of the institutional repositories are discipline-specific, too (e.g., marine research), GBIF.
- General use repositories. Zenodo is preferable since it is simple to connect with the EC reporting system, but others like Figshare are available, too. Particular attention has to be paid to the description/abstract text so that it's rich and descriptive. The downside of being a general-use repository is that the metadata schema must remain as agnostic as possible, and in many cases is insufficient for proper assessment of fitness-for-purpose.

There will be a list of OLAMUR datasets published and their identifiers available as described in section 9 and the EC reporting system.

5.1 Opting out of Open Access

The fulfillment of Open Access requirements for data in Horizon 2020 and subsequently in Horizon Europe is predicated in the general principle of "as open as possible, as closed as necessary". In OLAMUR some data must opt out of open access and remain restricted during and after the project lifetime. The reasons fall in two principal categories:

- Commercial exploitation and security of its operations. Detailed maps and data from OWFs that include the location of foundations and cables are provided by the commercial partners that run the wind farms (e.g., Vattenfall). They shall only be used as input to generate the micro-sitting maps (WP2) and not be made publicly available, unless the commercial partners decide to do so. In the latter case, they shall also decide the level of data granularity and detail the published dataset shall have.
- Personal data subjected to GDPR. Raw data from interviews (from WP7 and 2) cannot be shared, and shall be anonymized and stripped of personal data prior to publication. Personal data used for administration purposes (e.g., event organization) will be managed following GDPR guidelines (see section 7) and not be published, since it is not a relevant research output.

6. Data security

The data will be safely stored in both the originating data collecting system for partners and in international repositories for long term preservation and curation.

6.1 OLAMUR backend

"Funded by the European Union, grant agreement NO 101094065. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting authority can be held responsible for them."

In the OLAMUR backend, WP5 will not manage personal data, and the access to datasets in the private instances will be possible thanks to the creation of anonymous accounts that cannot be traced back to specific people or partners. The infrastructure is going to be deployed on the ARUBA.it infrastructure (https://business.aruba.it/azienda.aspx). Since 2015 Aruba is running a dedicated service to private business clients and it provides the client with top level services such as Data Centre (Virtual Servers, Real Servers, hosting infrastructures), Backup and Disaster Recovery etc. ARUBA also provides us with the most recent services for data security cryptography (AES), security protocols (AES, SSL) and bandwidth balance. The main characteristics of the service are summarized in Table 4:

SLA	99,80%
security	crypted transmission channel
(optional) storage crypting AES-256	
min backup timing	1h
schedule	anytime
granularity – single backup job	
Backup Account number	unlimited
concurrent agents	depends on the agreed service
max number of backup jobs	unlimited
bandwidth	unlimited (upload/download)
Certifications	ISO 9001:2015, ISO 27001:2013
service desk	24h
Cloud security certification	ISO/IEC 27017:2015
data privacy	ISO/IEC 2018:2014
security incidents:	ISO/IEC 27035:2016

Table 4: Security characteristics of the OLAMUR backend

7. Ethical aspects

WPs 7 and 9 will handle personal data, which requires particular attention. Data protection regulations in OLAMUR are in line with EU data protection policies following the GDPR. No transfer of personal data outside the EU and EEA areas is expected in OLAMUR.

To fully comply with art. 37 of GDPR "Designation of the Data Protection Officer" all partners which are public entities will be required to confirm that a Data Protection Officer (DPO) has been appointed in each organisation and the contact details of the DPO are made available to all data subjects involved in the research.

The following elements define approach on GDPR:

- List of DPO for each beneficiary (the list will be updated annually to keep track of changes)
- Explanation on how all the data to be processed are relevant and limited to the purpose of the project, in accordance with the data minimisation principle
- Description of the technical and organisational measures that will be implemented to

"Funded by the European Union, grant agreement NO 101094065. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting authority can be held responsible for them."

safeguard the rights of the data subjects/research participants

- Description of the security measures that will be implemented to prevent unauthorised access to personal data
- Description of the security measures that will be implemented to prevent unauthorised access to the equipment
- Description of the anonymisation/pseudonymisation techniques to be implemented
- Description of the management of personal data transfer from the EU to a non-EU country or international organisation

The Data Manager (ETT SPA), holds the responsibility for data management plan, strategy and the ways in which data used within OLAMUR are collected, processed, handled, distributed, protected, and stored. This applies to all research data collected during the project execution. On the other hand, the DPOs are responsible for the management of personal data used by their institution. This represents a further measure to ensure that personal data are not subject to any misuse. The DM interacts with DPOs to present and share OLAMUR recommendations on data protection. DPOs have to communicate to the DM their management strategy and the ways data (personal data, sensitive non-personal data, confidential data, personal and sensitive metadata) used within OLAMUR will be collected, processed, handled, distributed, protected and stored.

All websites of the project will also respect information technologies laws and civil rights and state privacy policies about cookies and personal data collection.

The DC will monitor the strict application of GDPR by all consortium members and everybody dealing with OLAMUR (personal) data.

Many Partners are managing DPO at institute level. In this case, DPO should add the project in their projects list. In case of changes, or new appointments, DMP will be updated accordingly.

Informed consent for data sharing and long-term preservation included in questionnaires dealing with personal data.

Whenever the OLAMUR project will implement surveys, questionnaires, or collect personal data for any reason (e.g., attendance to events), European GDPR regulation will be used as reference and the user will be informed about the use of personal data.

In general, OLAMUR will not transfer personal data (e.g., email addresses) to other entities and the only use will be setting up a distribution list to inform users about project progress. User will be always able to change their consensus and ask for being removed from the distribution channel.

For stakeholder engagement activities, coordinators will collect only the necessary information, including name and surname, affiliation, and activity sector/field of expertise.

Information will be collected by using information sheets and consent forms. Any further contact by the coordinator the activity should be explicitly authorised by the data subject, and it will be strictly limited to the purposes stated in the consent form.

"Funded by the European Union, grant agreement NO 101094065. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting authority can be held responsible for them."

Contact information will be used to involve stakeholders who gave their consent in follow up communications and efforts, in order to guarantee consistency with OLAMUR multi-stakeholder approach.

Workshops and dissemination events will be conducted according to two different approaches throughout the project timeframe, based on the purpose and scope of the initiative:

- Events without registration: for these events, coordinators will collect no data about participants, in line with the principle of data minimisation. This approach will be followed whenever it does not impair the fulfilment of OLAMUR objectives.
- Events with registration: in order to take part in these events (e.g., workshops), the participants will need to submit a registration form, with a limited number of personal data fields (up to 7). In this case, basic personal information and contact data will be used to facilitate the management of the meeting, give participants access to preparatory materials. Participants will have the possibility to give their consent to receive information on the organisation of further initiatives and related activities.

In accordance with the EU Charter of Fundamental Rights and with the GDPR, OLAMUR will safeguard the following rights throughout the project duration: (i) right to be informed about the personal information shared, (ii) right of access to such information, (iii) right to rectification (in case the information is incorrect or incomplete), (iv) right to erasure of the data, (v) right to restrict processing of the data, (vi) right to data portability from a service to the other, (vii) right to object.

Concerning the right to be informed (art. 13 and art. 14 GDPR) about the personal data shared with the Consortium (or with one of the partners), when giving the consent to data processing, data subjects will be adequately informed, and will be able to access information on data sharing on the portal of the activity for which data are collected.

Concerning the right to access information shared (art. 15 GDPR), data subjects will have the possibility to request a copy of such data to the relevant DPO. Based on the internal procedures of the DPO, the data stored by the organisation will be collected within 30 days. Any potential delay or obstacle (due to the anonymisation/pseudonymisation procedure or to the data storage techniques) will be adequately notified and justified to the data subject.

Similarly, concerning the right to rectify incorrect or incomplete information (art.16 GDPR), OLAMUR will ensure that data subject have the possibility to submit a request of rectification to the relevant DPO, that will carry out the necessary procedures within 30 days. Any potential delay will be adequately notified and justified to the data subject.

Concerning the right to be forgotten (art. 17 GDPR), OLAMUR will ensure the erasure of the data collected and of the files stored in local archives within 30 days from the official request. When sharing personal data, the subject will be informed of the procedures to obtain the erasure of such data, according to the internal policies and data management infrastructures of the partner leading a certain activity.

"Funded by the European Union, grant agreement NO 101094065. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting authority can be held responsible for them."

Concerning the right to portability of data and their limitation (art. 18 and art 20 of GDPR), the subject will be informed on the procedures to obtain the transmission of data to the physical or legal person designated, based on the internal policies of the responsible partner. The partner will be responsible for the assessment and management of the request and for the aggregation of data in an intelligible format in due time.

8. Datasets information form

As a previous step to the creation of this document, a dataset information form was circulated to the project participants, who were requested to fill it by WP. It provides a first overview of the data that will be both used and generated in OLAMUR and forms one of the basis for this document and the development of the OLAMUR backend. It is hosted in the internal Sharepoint space for the project participants.

The template was developed based on a form created in the MEESO project (Soni *et al.*, 2020) which followed the EC Guidelines (Guidelines on FAIR Data Management in Horizon 2020, 2016). It was further adapted to the OLAMUR project and example responses were added to aid the researchers. Table 5 summarizes the questions included in the template.

Partner	*Institutional partner*
WP	
Contact person and email	*Of the person responsible for the dataset*
DATA SUMMARY	
Dataset name / brief description	
Preexisting data or generated in OLAMUR? (dropdown menu)	*Generated in OLAMUR / Pre-existing data / Generated from pre-existing data*
If already existing, DOI or link	
Purpose, target user	*State the purpose of the data collection/generation, and which other WPs will use it, if applicable*
Observation or model?	
Expected size	
Level of processing and description	*Describe if it's raw, calibrated, postprocessed… Describe calculations, averaging, interpolation performed. QC procedures, etc…*
Variables and units	
Spatial extent and resolution	*Resolution particularly important for gridded and map data*

"Funded by the European Union, grant agreement NO 101094065. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting authority can be held responsible for them."

Time span and resolution	
Metadata provided	*If some standard is followed, specify*
TIMING	
When will be ready to share with partners	*By month*
When will be ready to publish	*Made publicly available. If applicable*
INTEROPERABILITY	
File format	
Will you use standard vocabularies? Which? (see info tab)	
If you need to your own vocabulary, will you provide mappings?	
ACCESSIBILITY	
Will the data be open access?	
In it can't be published, provide the reason	*Specify the reasons: ethical rules of personal data, intellectual property, commercial, privacy-related, security-related*
Which repository will you use?	
Are there any ethical or legal concerns?	
Which license will you use?	

Table 5: Template for the preliminary description of OLAMUR data to be generated and/or collected during the project.

9. List of available datasets

The list of publicly available OLAMUR datasets with their corresponding PIDs will be added to this document as they are generated and this document updated. A more up-to-date list will be available in the OLAMUR website https://olamur.eu/.

10. References

European Commission. (2016). Guidelines on FAIR Data Management in Horizon 2020 v3.0. (Directorate-General for Research and Innovation) p. 12. Retrieved on 31-05-2018, from http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf

"Funded by the European Union, grant agreement NO 101094065. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting authority can be held responsible for them."

FORCE11 Data Citation Synthesis Group (2014) Joint Declaration of Data Citation Principles. Martone M. (ed.). https://doi.org/10.25490/a97f-egyk

Mons, B., Neylon, C., Velterop, J., Dumontier, M., da Silva Santos, L. O. B., & Wilkinson, M. D. (2017). Cloudy, increasingly FAIR; revisiting the FAIR Data guiding principles for the European Open Science Cloud. Information services & use, 37(1), 49-56. https://doi.org/10.3233/ISU-170824

Simons, R.A., and Chris John. 2022. ERDDAP. https://coastwatch.pfeg.noaa.gov/erddap. Monterey, CA: NOAA/NMFS/SWFSC/ERD.

Soni, V., Cronne-Grigorov, L., Holdsworh, N. (2020). Data Management Plan in cooperation with WP1. MEESO project. https://ec.europa.eu/research/participants/documents/downloadPublic?documentIds=080166e5ceac89ef&appId=PPGMS

Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., ... & Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. Scientific data, 3(1), 1-9. https://doi.org/10.1038/sdata.2016.18

"Funded by the European Union, grant agreement NO 101094065. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting authority can be held responsible for them."