



Project acronym: **GROOM II**

Project title: **G**liders for **R**esearch, **O**cean **O**bservation & **M**anagement:
Infrastructure and **I**nnovation

Grant agreement no. 951842

D.1.2

Data Management Plan – Initial DMP

Due delivery date: M6

Organization name of lead participant for this deliverable: ARMINES

Dissemination level		
PU	Public	
CO	Confidential, only for members of the consortium	X
Deliverable number	D.1.2	
Deliverable responsible	ARMINES	
Work Package	WP1 – Project Management and Conceptual Design Reporting	



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 951842

Author(s) – in alphabetical order	
Laurent Mortier	ENSTA
Kamil Szafranski	ARMINES

Document revision history			
Version	Date	Modifications introduced	
		Change reason	Modified by
v0	07/2021	Outline of the deliverable	Kamil Szafranski, ARMINES
v1	08/2021	1st version of the deliverable	Laurent Mortier, ENSTA
v2	09/2021	2nd version of the deliverable	Kamil Szafranski, ARMINES Daniel Hayes, CSCS João Sousa, UPORTO
final	10/2021	Final review	Kamil Szafranski, ARMINES

Deliverable abstract
<p>The Data Management Plan (DMP) is a public deliverable which corresponds to the deliverable D1.2, produced in the context of GROOM II project WP1 – Project Management and Conceptual Design Reporting.</p> <p>The present DMP defines a common strategy for issues such as sharing and archiving of data generated by the GROOM II project, the length of its curation and preservation etc. This document defines the data management plan that will be followed in order to improve and maximize access to, storage and re-use of the data generated by the project.</p>



Table of contents

1. Background and context	4
2. Data Summary	5
2.1. PURPOSE OF THE DATA COLLECTION/GENERATION	5
2.2 RELATION TO THE OBJECTIVES OF THE PROJECT	5
2.3 TYPES AND FORMATS OF DATA GENERATED/COLLECTED	5
2.4 DATA UTILITY AND RE-USE	6
3. FAIR data	7
3.1 MAKING DATA FINDABLE, INCLUDING PROVISIONS FOR METADATA:	7
3.2 MAKING DATA OPENLY ACCESSIBLE:	7
3.3 MAKING DATA INTEROPERABLE:	8
3.4 INCREASE DATA RE-USE (THROUGH CLARIFYING LICENSES):	8
4. Allocation of resources	9
4.1 COST ESTIMATION FOR MAKING GLIDER DATA (GD) DATA FAIR	9
4.2 CLEARLY IDENTIFY RESPONSIBILITIES FOR DATA MANAGEMENT IN YOUR PROJECT	9
4.3 COSTS AND POTENTIAL VALUE OF LONG TERM PRESERVATION	9
5. Data security	10
6. Ethical aspects	10
7. Other	10



List of Abbreviations

Argo	Scientific international programme for ocean observation using a fleet of robots
ASV	Autonomous Surface Vehicle
AUV	Autonomous Underwater Vehicle
EC	European Commission
EECP	European Cluster Collaboration Platform
EMBRC	European Marine Biological Resource Centre
EMODnet	European Marine Observation and Data Network
EMSO	European Multidisciplinary Seafloor and water column Observatory
EOOS	European Ocean Observing System
EuroArgo	European contribution to the Argo Programme
GERI	Glider European Research Infrastructure
GOOS	Global Ocean Observing System
IOC	Intergovernmental Oceanographic Commission
IMOS	Integrated Marine Observing System
IOOS	Integrated Ocean Observing System
JCOMM	Joint Technical Commission for Oceanography and Marine Meteorology
JERICO	Joint European Research Infrastructure of Coastal Observatories: Science, Service, Sustainability
MRI	Marine Research Institute
MS	Member States
OCG	Observations Coordination Group
R&D	Research & Development
SME	Small and Medium Enterprise
WP	Work Package

DISCLAIMER

The contents of this publication are the sole responsibility of ARMINES and do not necessarily reflect the opinion of the European Union.



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 951842

1. Background and context

Underwater and surface autonomous vehicles, in particular gliders, have become essential for many environmental observations from the surface down to 6000 m and for activities supporting the blue economy. Their major advantages are being mobile, steerable, persistent and usable in large numbers and at relatively low costs with a variety of scientific payloads available. However, the distributed infrastructure required to exploit these assets must be able to meet different demands from research and monitoring of the marine environment to public service missions and industry needs, requiring customised payloads and operations. The rapid evolution of such technologies (robotics, artificial intelligence, sensors, big data) requires that the R&D resources offered by this distributed infrastructure continuously adapt to users' demands.

The complex hardware and information technology characteristics of such a distributed European infrastructure, optimizing access to resources and R&D for gliders, were analysed during the GROOM-FP7 design study from the perspective of research and the Global and (future) European Ocean Observing System (GOOS & EOOS) needs. Since then, several “gliderports” have been developed which has fostered a corresponding European industrial innovative sector.

GROOM II, building on its predecessor, will deliver the decision basis for an advanced Marine Research Institute (MRI) that promotes scientific excellence, fosters innovation, supports the blue economy, builds industrial and public partnerships, and works towards helping achieve the common research and innovation mission for future Europe. The project will define the overall organization of an infrastructure dedicated to ocean research and innovation, and maritime services supporting Blue Growth: The Glider European Research Infrastructure (GERI).

This infrastructure will be a positive step against today's fragmented European landscape, aiding connections, and synergies for the completion of GOOS and EOOS.

2. Data Summary

Provide a summary of the data addressing the following issues:

State the purpose of the data collection/generation

Explain the relation to the objectives of the project

Specify the types and formats of data generated/collected

Specify if existing data is being re-used (if any)

Specify the origin of the data

State the expected size of the data (if known)

Outline the data utility: to whom will it be useful

2.1. PURPOSE OF THE DATA COLLECTION/GENERATION

The Data Management Plan (DMP) is a public deliverable which corresponds to the deliverable D1.2, produced in the context of GROOM II project WP1 – Project Management and Conceptual Design Reporting.

The GROOM II consortium participates in the “Pilot on Open Research Data in Horizon 2020” because of the shared goal to foster access to research data, which happens to be generated in H2020 project, by enabling open access and reuse. The development of a Data Management Plan is also required for all projects participating in the Open Research Data Pilot.

2.2 RELATION TO THE OBJECTIVES OF THE PROJECT

The present DMP defines a common strategy for issues such as sharing and archiving of data generated by the GROOM II project, the length of its curation and preservation etc. This document defines the data management plan that will be followed in order to improve and maximize access, storage and reuse of the data generated by the project.

2.3 TYPES AND FORMATS OF DATA GENERATED/COLLECTED

The following data sets are foreseen:

- Glider and similar platforms scientific and engineering Data (GD)
- Test Data (TD)
- Commercial Data (CD)
- Budgetary and Financial Data (BFD)

- Product Data (PD)
- Marketing Data (MD)
- Personal Data (PsD)

The majority of the data produced will be numerical data, with potentially some image files. These data will be stored on the computers associated with the equipment where they were generated. All partners have backup servers, where copies of the data will be stored in a first instance. Personnel working on the project will be permitted to have copies of the data on personal computers supplied by the partners.

The treatment of personal data (PsD) will fulfil the requirements from the European Commission in terms of ethics and protection of personal data (POPD) as detailed in the GROOM II Deliverable 7.1 "POPD - Requirement No.1".

2.4 DATA UTILITY AND RE-USE

The data collected and generated by the project concerns only data and metadata on the existing infrastructures maintained or planned by the partners of the project and its stakeholders. The GROOM II partners will not generate for the project any geophysical and biogeochemical marine data (or glider data GD) that already have well defined data management methodology and policies on which to rely (perhaps thanks to global, European and national organizations or projects). However, the DMP considers these GD for the sake of completeness since the data management plan of the GD that will be collected by the designed infrastructure is an important deliverable of the project itself. The present DMP for the GD does not prejudice the future DMP.

Except for glider data (GD) collection at sea carried out during the project by the partners, the data produced is not anticipated to be the type of (meta) data of interest to outside parties. Test Data (TD), Commercial Data (CD), Budgetary and Financial Data (BFD), Product Data (PD) and Marketing Data (MD) will be accessible by all partners on cloud storage to be used from M0 to M36. Most of these data will be used in aggregated form for the public deliverables and other similar publishable documents of the project, including websites. The distribution of these data in raw format is not foreseen, but may be considered upon request after a disclosure agreement by the partners according to the consortium agreement.

Glider data (GD) collected at sea as part of usual scientific, statutory monitoring or commercial activities of the partners, will be disseminated to the public in near real time with a specific quality control following the metadata and data flow architecture designed during GROOM-FP7 and since developed further. GD will be published in delayed mode by the Principal Investigators (PIs), after a 'delayed mode' quality check thanks to the same architecture, with a recommended delay of 6 months after the data collection. The recommendations for data formatting and publication will ensure FAIR data publication.

3. FAIR data

3.1 MAKING DATA FINDABLE, INCLUDING PROVISIONS FOR METADATA:

Outline the discoverability of data (metadata provision)

Outline the identifiability of data and refer to standard identification mechanisms. Do you make use of persistent and unique identifiers such as Digital Object Identifiers?

Outline naming conventions used

Outline the approach towards search keyword

Outline the approach for clear versioning

Specify standards for metadata creation (if any). If there are no standards in your discipline describe what metadata will be created and how

The Everyone's Gliding Observatories (EGO) initiative is a gathering of several teams of oceanographers around the world who are interested in developing the use of gliders and similar platforms for ocean observations. EGO-recommended best practices for data-management encourages the storage and sharing of (meta-) data regarding:

- the platform (main characteristics of the glider),
- the deployment (main characteristics of the glider deployment),
- the instrument,
- and the sensors used,

through a data management system similar to the Argo one (designed for profiling floats), and under similar data and metadata format and technical data standards. Such standards are endorsed by international organizations and programs, such as OceanGliders, the associated GOOS program for glider observations, which have developed in the past decade a strong corpus of standards, methodologies and tools that have been widely adopted by the glider community.

3.2 MAKING DATA OPENLY ACCESSIBLE:

Specify which data will be made openly available? If some data is kept closed provide rationale for doing so

Specify how the data will be made available

Specify what methods or software tools are needed to access the data? Is documentation about the software needed to access the data included? Is it possible to include the relevant software (e.g. in open source code)?

Specify where the data and associated metadata, documentation and code are deposited

Specify how access will be provided in case there are any restrictions

Glider data (GD) collected at sea will be disseminated to the public with a specific quality control following the metadata and data flow architecture designed during GROOM-FP7 and further developed. GD are published in delayed mode by the Principal Investigators (PIs), after a 'delayed mode' quality check thanks to the same architecture, with a recommended delay of 6 months after the data collection. The recommendations for data formatting and publication will ensure FAIR data publication.

Other Material produced by the partners during the project will be published on repositories with proper metadata and/or licence. Open source software will be encouraged in the framework of GROOM II. Tools such as GitHub are also encouraged to publish code and documentation, which only requires internet access, a web browser and a free registration. Methods related to data management and collection will be published in open access platforms such as the Ocean Best Practices Repository.

3.3 MAKING DATA INTEROPERABLE:

Assess the interoperability of your data. Specify what data and metadata vocabularies, standards or methodologies you will follow to facilitate interoperability.

Specify whether you will be using standard vocabulary for all data types present in your data set, to allow inter-disciplinary interoperability? If not, will you provide mapping to more commonly used ontologies?

Glider data (GD) and metadata containing oceanographic time series will be encoded under EGO glider common format and archived in global data centres (Coriolis). Once confirmed, the new common format designed by OceanGliders (OG1.0) will be encouraged.

The EGO glider data received in real-time are quality-controlled. The real-time quality control procedures are described in the EGO glider quality control manual. They are automatically applied, without human intervention to minimize the delay between data observation and data distribution. For each active glider, the data that passed the real-time QC tests are distributed on GTS (the WMO data transmission system). Whenever possible, glider mission metadata will be registered on the OceanGliders Dashboard at OceanOps (<https://www.ocean-ops.org/board?t=oceangliders>).

3.4 INCREASE DATA RE-USE (THROUGH CLARIFYING LICENSES):

Specify how the data will be licensed to permit the widest reuse possible

Specify when the data will be made available for reuse. If applicable, specify why and for what period a data embargo is needed

Specify whether the data produced and/or used in the project is usable by third parties, in particular after the end of the project? If the re-use of some data is restricted, explain why

Describe data quality assurance processes

Specify the length of time for which the data will remain re-usable

Gliders data (GD) will be accessible to public in delayed (6 months) mode and will be archived and made available for re-use in Coriolis data service, providing quality-controlled *in situ* data in real-time and delayed modes (<http://www.coriolis.eu.org/Data-Products/Data-Delivery/Data-selection>). Managed data sets are mainly T-S profiles and time series from gliders, profiling floats, XBT's, thermo-salinographs, drifting and moored buoys. Coriolis is progressively integrating other parameters such as sea level from European real-time tide gages, and ecosystem data (oxygen, chlorophyll and nutrients) from ferrybox, mooring and gliders.

Transmitted data are processed, checked for quality and distributed to the GTS and Internet. The data service works in close association with a scientific team to define procedures for data validation, quality control, formats and products.

4. Allocation of resources

Explain the allocation of resources, addressing the following issues:

Estimate the costs for making your data FAIR. Describe how you intend to cover these costs

Clearly identify responsibilities for data management in your project

Describe costs and potential value of long term preservation

4.1 COST ESTIMATION FOR MAKING GLIDER DATA (GD) DATA FAIR

No additional costs required

4.2 CLEARLY IDENTIFY RESPONSIBILITIES FOR DATA MANAGEMENT IN YOUR PROJECT

Project data and technical director responsible for data management will be hired by the end of 2021.

4.3 COSTS AND POTENTIAL VALUE OF LONG TERM PRESERVATION

No additional costs required

5. Data security

Address data recovery as well as secure storage and transfer of sensitive data

There are two GDACs (Global Data Assembly Centers) for redundancy, which are the users' access points for EGO data. One GDAC is located in France (Coriolis, <http://www.coriolis.eu.org>), the other one in the USA (NDBC, National Data Buoy Center, <http://www.ndbc.noaa.gov>). The GDACs handle EGO data, metadata, and index files on ftp servers. The servers at both GDACs are synchronized at least daily to provide the same EGO data. The GDAC distributes the best copy of the data files. When a higher quality data file (e.g. calibrated data) is available, it replaces the previous version of the data file.

No sensitive data will be transferred.

6. Ethical aspects

To be covered in the context of the ethics review, ethics section of DoA and ethics deliverables. Include references and related technical aspects if not covered by the former

Whenever the GROOM II project will implement surveys, questionnaires, or collect personal data for any reason (e.g., attendance to organized events), European GRDP law will be used as reference and the user will be informed about the use of personal data. In general, GROOM II 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 his consensus and ask for it to be removed from the distribution channel.

In case personal data are transferred from the EU to a non-EU country or international organisation, confirmation that such transfers are in accordance with Chapter V of the General Data Protection Regulation 2016/679, will be submitted as a deliverable. In case personal data are transferred from a non-EU country to the EU (or another third state), confirmation that such transfers comply with the laws of the country in which the data was collected will be submitted as a deliverable too.

Ethics and protection of personal data (POPD) as detailed in the GROOM II Deliverable 7.1 "POPD - Requirement No.1".

7. Other

Refer to other national/funder/sectorial/departmental procedures for data management that you are using (if any)

Nothing to report