



SeaDataNet data
management protocols
for glider data

WP9 – Deliverable D9.14



HORIZON 2020

sdn-userdesk@seadatanet.org – www.seadatanet.org

SeaDataCloud - Further developing the pan-European infrastructure for marine and ocean data management

Grant Agreement Number: 730960

Deliverable number

D9.14

Short title

SDN protocols for glider data

Long title

SeaDataNet data management protocols for glider data

Short description

This document presents data managers with an overview of global glider data management best practice and outlines recommendations for assimilation of glider observations into the European SeaDataNet infrastructure.

Author

M. Hebden, J. Buck

Working group

WP9

Dissemination

Public

Copyright terms

History

| Version | Authors | Date | Comments |
|---------|--------------------|------------|--|
| 1.0 | M. Hebden, J. Buck | 28/08/2019 | Description of protocols for glider data within SeaDataNet |
| 1.1 | M. Hebden | 12/09/2019 | Incorporation of feedback from Antonio Novellino, ETT |
| 1.2 | D.M.A. Schaap | 13/10/2019 | Review and acceptance by SDC Technical Coordinator |
| | | | |
| | | | |



sdn-userdesk@seadatanet.org – www.seadatanet.org

SeaDataCloud - Further developing the pan-European infrastructure for marine and ocean data management

Grant Agreement Number: 730960

Disclaimer

The content of this document reflects only the authors' view; it cannot be considered to reflect the view of the European Commission or any other body of the European Union. The European Commission is not responsible for any use that may be made of the information it contains.

Acknowledgment

This document is a deliverable of the SeaDataCloud project. This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 730960.



sdn-userdesk@seadatanet.org – www.seadatanet.org

SeaDataCloud - Further developing the pan-European infrastructure for marine and ocean data management

Grant Agreement Number: 730960

Table of contents

| | |
|--|----|
| 1. Introduction | 6 |
| 1.1. Identifying the key stakeholders | 6 |
| 1.1.1. Glider operators | 6 |
| 1.1.2. Glider data managers | 6 |
| 1.1.3. Everyone’s Gliding Observatories (EGO) | 6 |
| 1.1.4. The Intergovernmental Oceanographic Commission of the United Nations Educational, Scientific and Cultural Organization (IOC-UNESCO) | 7 |
| 1.1.5. Global Ocean Observing System (GOOS), European Ocean Observing System (EuroGOOS) and the OceanGliders Program. | 7 |
| 1.1.6. The Copernicus Marine Environment Monitoring Service (CMEMS) | 7 |
| 1.1.7. SeaDataNet infrastructure | 8 |
| 1.1.8. European Marine Observation and Data network (EMODnet) | 8 |
| 1.1.9. The JCOMM in-situ Observations Programme Support Centre (JCOMMOPS)..... | 9 |
| 1.1.10. International Council for the Exploration of the Sea (ICES) | 9 |
| 1.1.11. World Meteorological Organization (WMO) and Global Telecommunication System (GTS) | 9 |
| 1.1.12. The U.S Integrated Ocean Observing System (IOOS) | 10 |
| 1.1.13. The Australian Integrated Marine Observing System (IMOS) and the Australian National Facility for Ocean Gliders (ANFOG)..... | 10 |
| 1.2. Fundamental principles underpinning glider data management | 10 |
| 1.3. Goals of deliverable and structure for remainder of report | 11 |
| 2. A brief history of glider data management | 11 |
| 2.1. Europe: EGO to GROOM to OceanGliders to GOOS | 11 |
| 2.2. Achieving global interoperability | 13 |
| 3. Glider data management principles and best practice | 13 |
| 3.1. Data versioning | 13 |
| 3.1.1. Near Real Time (NRT) data stream..... | 13 |
| 3.1.2. Recovery (REC) data stream..... | 13 |
| 3.1.3. Delayed Mode (DM) data stream..... | 13 |
| 3.2. Data (and metadata) standardisation | 14 |
| 3.2.1. Data exchange formats | 14 |
| 3.2.2. Controlled vocabularies | 15 |
| 3.3. Data Quality Control (QC) and Quality Assurance (QA) | 16 |
| 3.4. Data exchange pathways and tools..... | 17 |
| 3.4.1. The Global Data Assembly Centre (GDAC) model | 17 |
| 3.4.2. GTS (Global Telecommunication System) data assimilation | 19 |
| 3.4.3. THREDDS and ERDDAP | 19 |
| 3.4.4. Open Geospatial Consortium (OGC) Standards and Sensor Web Enablement (SWE) | 20 |
| 3.5. Common community tools for processing and quality control of glider data | 20 |
| 4. Integration with SeaDataNet | 21 |
| 4.1. Rationale for integration | 21 |
| 4.2. AtlantOS recommendations for integration..... | 21 |
| 4.2.1. Implementation of AtlantOS recommendations for Argo..... | 21 |
| 4.3. Mapping the EGO data exchange format (V1.2) to CDI | 23 |



| | |
|--|----|
| 4.3.1. Key observations from CDI-EGO mapping..... | 26 |
| 4.4. Near real time pathways of glider observations into SeaDataNet and related European infrastructures | 27 |
| 4.4.1. OGC-SWE route: The ‘shop window’ approach | 27 |
| 4.4.2. Assimilation of NRT data collections | 28 |
| 4.5. Data dissemination by SeaDataNet..... | 28 |
| 4.6. Summary of recommendations and concluding remarks | 29 |
| 5. References..... | 31 |
| 6. List of Acronyms | 32 |



sdn-userdesk@seadatanet.org – www.seadatanet.org

SeaDataCloud - Further developing the pan-European infrastructure for marine and ocean data management

Grant Agreement Number: 730960

1. Introduction

The ocean glider (referred to hereafter as ‘glider’) has manifested itself as a ‘go to’ platform for cost-effective and increasingly reliable acquisition of near real time oceanographic data.

Modern day profiling glider platforms include the Slocum, Seaglider, SEAEXPLORER and Spray models, all utilising the basic Autonomous Underwater Vehicle (AUV) design pioneered by Doug Webb and Henry Stommel in the 1980s.

The glider is a versatile platform capable of carrying a payload of increasingly diverse oceanographic sensors, with information harvested either from internal data storage on-board, or via near real time telemetry bursts to a base station on shore.

The glider platform is used in a variety of ways, ranging from discrete one-off scientific missions (or ‘process studies’) to more strategic missions, often as components of long-term operational observatories.

There are a whole host of end users of glider data including academia, industry, military and government policy makers.

The purpose of this document is to: i) ensure that SeaDataNet partners exposed to glider data are aware of glider data management best practice and have the knowledge needed to assimilate glider data into SeaDataNet and ii) advise SeaDataNet itself on how best to align with and help steer emerging international protocols for glider data management and exchange.

1.1. Identifying the key stakeholders

In this section we identify some of the principal global stakeholders working with gliders and glider data. These stakeholders are responsible for ensuring the flow of quality glider observations to a wide range of data end users, such as policy makers, scientists and climate forecasters. An appreciation of the range of stakeholders involved in glider data management is important before attempting to introduce new data flows (e.g. SeaDataNet integration).

1.1.1. Glider operators

Scientists and technicians using glider technology to observe the ocean, who may have a desire to make collected data available for reuse, thereby adding greater impact to their work.

1.1.2. Glider data managers

Individuals or teams responsible for collating standalone, interoperable information from gliders who have a vested interest in adopting community best practice for glider data management. This includes National Oceanographic Data Centres (NODCs) and glider Data Assembly Centres (DACs).

1.1.3. Everyone’s Gliding Observatories (EGO)

Everyone’s Gliding Observatories (EGO) is an open community initiative that brings together glider operators, manufacturers, data managers and policy makers. The EGO acronym initially stood for European Gliding Observatories, but ultimately changed to reflect increasing involvement of non-European collaborators. EGO provides a focal point to present, discuss and resolve scientific and



sdn-userdesk@seadatanet.org – www.seadatanet.org

SeaDataCloud - Further developing the pan-European infrastructure for marine and ocean data management

Grant Agreement Number: 730960

technological issues around glider operations and helps to drive and formalise associated data management protocols, including data exchange. The first EGO community meeting took place in 2005, with subsequent gatherings taking place every 1-3 years thereafter.

Since its inception, the EGO initiative has established an infrastructure for glider data exchange, based around national Data Assembly Centres (DACs) – nodes for feeding open access, interoperable data and metadata from gliders to a central Global Data Assembly Centre (GDAC), based at Coriolis, France.

1.1.4. The Intergovernmental Oceanographic Commission of the United Nations Educational, Scientific and Cultural Organization (IOC-UNESCO)

IOC-UNESCO was created in 1960 as a body with functional autonomy within UNESCO. The Commission functions as the authoritative voice on marine science within the United Nations (UN) system, promoting international cooperation for research, services, management and sustainable development of the marine environment.

1.1.5. Global Ocean Observing System (GOOS), European Ocean Observing System (EuroGOOS) and the OceanGliders Program.

The Global Ocean Observing System (GOOS) is a worldwide infrastructure established to ensure the sustained flow of standardised, interoperable ocean measurements from in situ networks, satellite observations, governments, UN agencies and individual scientists. GOOS was first set up in 1991 by IOC-UNESCO, with other sponsors, as a platform for global cooperation and information exchange, ensuring current and future sustainability of the global oceans.

The European Global Ocean Observing System (EuroGOOS) is the European component of GOOS, founded in 1994. It is one of thirteen GOOS Regional Alliances (GRAs). EuroGOOS serves 44 member organisations and supports five connected Regional Operational Oceanographic Systems (ROOS). A function of EuroGOOS is the implementation of the European Ocean Observing System (EOOS) framework, which aims to align and integrate Europe's ocean observing capability.

Under EuroGOOS a Glider Task Team, with the support of EMODnet Physics (and composed of European DACs and the GDAC) was established in 2015. This is a consortium to champion and progress glider data management best practice and data flow within Europe. This group is essentially a subset of the international OceanGliders Data Management Task Team (OGDMTT) established during 2016 under GOOS to coordinate international glider data management (and integration with GOOS).

1.1.6. The Copernicus Marine Environment Monitoring Service (CMEMS)

The Copernicus Marine Environment Monitoring Service (CMEMS) utilises information from satellites and in situ observations to provide regular, reliable and free open access to reference information on the state of the global ocean and European regional seas. This core marine service targets users within four main sectors: i) maritime safety, ii) marine resources, iii) coastal and marine environment and iv) weather, seasonal forecasting and climate.



sdn-userdesk@seadatanet.org – www.seadatanet.org

SeaDataCloud - Further developing the pan-European infrastructure for marine and ocean data management

Grant Agreement Number: 730960

Near real time in situ measurements, such as those from glider deployments, are important to CMEMS as they present an opportunity to calibrate and validate the data from satellites, as well as model outputs, core components of CMEMS.

1.1.7. SeaDataNet infrastructure

SeaDataNet is a major European infrastructure providing access to interoperable ocean and marine datasets and products. These are acquired by European organisations from research cruises and other observational activities in the European coastal marine waters, regional seas and the global ocean. SeaDataNet develops, promotes and governs common marine standards for metadata and data formats, common vocabularies and quality flags as well as standard software tools and services. The core partners in SeaDataNet are the National Oceanographic Data Centres (NODCs) and major marine research institutes in Europe, together with the Intergovernmental Oceanographic Commission (IOC) of UNESCO and the International Council for the Exploration of the Sea (ICES). The SeaDataNet Data Centres are highly skilled and have been actively engaged in marine data management for many decades and have the essential capabilities and facilities for data quality control, long term stewardship, retrieval and distribution. SeaDataNet works together with originators of marine data, comprising scientists collecting data with the European research vessel fleet and other observing platforms, and various governmental agencies, collecting data for environmental management and economic activities. SeaDataNet also has a close cooperation with the European operational oceanography community.

Enhancement of the existing SeaDataNet infrastructure is currently underway, funded by the EU HORIZON 2020 SeaDataCloud project. A key goal of SeaDataCloud is adoption of cloud services and integration of High Performance Computing (HPC) technology, helping to realise the European Open Science Cloud (EOSC).

The SeaDataNet infrastructure underpins several portals of the European Marine Observation and Data network (EMODnet).

1.1.8. European Marine Observation and Data network (EMODnet)

The European Marine Observation and Data network (**EMODnet**), launched in 2008, is a long-term initiative of the European Commission Directorate General for Maritime Affairs and Fisheries (DG MARE), underpinning its Marine Knowledge 2020 strategy and supporting the marine data and information needs of the Integrated Maritime Policy (2007), the European Marine and Maritime Research Strategy (2008), the Marine Strategy Framework Directive (2008), and other actors in the marine and maritime economical communities.

Seven discipline-based themes are currently in place under the overall EMODnet umbrella (Bathymetry, Geology, Seabed habitats, Chemistry, Biology, Physics and Human Activities), several of which are underpinned by SeaDataNet standards, data and services. Each thematic group has its own gateway to bespoke services and open access data and data products. In addition, a central portal provides an overview of EMODnet activities and a product catalogue.

EMODnet Physics is a key glider stakeholder. EMODnet Physics is built on the thematic networks within EuroGOOS ROOSs (Regional Operational Oceanographic Systems), CMEMS INSTAC (Copernicus Marine Environment Monitoring Service In Situ Thematic Assembly Centre) and SeaDataNet-NODCs (SeaDataNet network of National Oceanographic Data Centres), and helps to bridge the gap between the near real time data streams (EuroGOOS ROOS and CMEMS INSTAC) and the archived, validated data managed within the SeaDataNet infrastructure.



sdn-userdesk@seadatanet.org – www.seadatanet.org

SeaDataCloud - Further developing the pan-European infrastructure for marine and ocean data management

Grant Agreement Number: 730960

A relatively recent addition to the EMODnet infrastructure has been the introduction of the EMODnet Data Ingestion portal, which is helping to mobilise additional near real time data streams within Europe, including those from emerging observational platform types, such as gliders. EMODnet Data Ingestion is helping to unlock data by means of interoperable machine-to-machine techniques (Open Geospatial Consortium-OGC, Sensor Web Enablement-SWE) and standards (ISO, NetCDF, IODE), in particular. These, coupled with tools and services offered by EMODnet Physics, are helping to promote greater discoverability, accessibility and usability of data streams by the wider community.

1.1.9. The JCOMM in-situ Observations Programme Support Centre (JCOMMOPS)

JCOMMOPS (Joint IOC/WMO Technical Commission for Oceanography and Marine Meteorology in situ Observing Platform Support Centre) facilitates the operational flow of metadata and data for existing operational ocean observing networks under JCOMM (and thus GOOS) and, as such, supports the sustainability of the OceanGliders Network. JCOMMOPS provides the practicalities needed to bind operational systems including metadata registry, technical guidance and monitoring tools to measure the health of operational information exchange infrastructures.

1.1.10. International Council for the Exploration of the Sea (ICES)

The International Council for the Exploration of the Sea (ICES) coordinates and promotes marine research on oceanography, the marine environment, the marine ecosystem and on living marine resources in the North Atlantic. ICES has a well-established Data Centre, which manages a number of large dataset collections related to the marine environment. The majority of the data – covering the North East Atlantic, Baltic Sea, Greenland Sea and Norwegian Sea – originate from national institutes that are part of the ICES network. The ICES Data Centre also provides marine data services to ICES member countries, expert groups, world data centres, regional seas conventions (HELCOM and OSPAR), the European Environment Agency (EEA), Eurostat and various other European projects and biodiversity portals.

ICES maintain the Platform Code (C17) vocabulary list, which has been adopted by EGO to formalise glider platform identification.

1.1.11. World Meteorological Organization (WMO) and Global Telecommunication System (GTS)

The World Meteorological Organization (WMO), a specialised agency of the UN, provides information on the state and behaviour of the Earth's atmosphere and associated interactions with land and oceans. A key mandate of the WMO is to facilitate international cooperation to enable international flow of weather and climate information.

WMO operates the Global Telecommunication System (GTS), an integrated network of surface and satellite-based telecommunication links connecting up world authorities in meteorological forecasting and climate analysis. The GTS enables the near real time exchange of data and information critical to these services.

Some of the core water column measurements routinely acquired from glider observations, such as temperature and salinity, are of great value to the WMO, helping to add integrity to weather forecasting models.



sdn-userdesk@seadatanet.org – www.seadatanet.org

SeaDataCloud - Further developing the pan-European infrastructure for marine and ocean data management

Grant Agreement Number: 730960

1.1.12. The U.S Integrated Ocean Observing System (IOOS)

The U.S Integrated Ocean Observing System (IOOS) operates a multi-platform integrated system to collect and disseminate information about the coastal waters, Great Lakes and oceans of national interest and to enable improved forecasting. The observing system incorporates in situ oceanographic instrumentation, coupled with land-based measurements and satellite remote sensing, underpinned by a centralised database to facilitate data flow to end users.

Glider observations are an integral part of the IOOS, providing valuable in situ near real time measurements. The IOOS Program Office invests heavily in glider technology, operating a glider DAC and has also established the Underwater Glider User Group (UG2).

1.1.13. The Australian Integrated Marine Observing System (IMOS) and the Australian National Facility for Ocean Gliders (ANFOG)

The Australian Integrated Marine Observing System (IMOS) is a national collaborative research infrastructure, supported by the Australian Government, with routine, fully integrated observation of the coastal and open ocean around the Australian continent. Data are made available to the marine and climatological community, including international collaborators.

IMOS operates a number of facilities as part of its network, including one for ocean glider operations – the Australian National Facility for Ocean Gliders (ANFOG). ANFOG operates the Australian glider fleet and disseminates data via the national Australian Ocean Data Network (AODN).

1.2. Fundamental principles underpinning glider data management

Glider technologies are still evolving. As such, associated data management infrastructures and best practice are still reaching optimum maturity. Similar, more established, communities, such as Argo, have helped to steer and shape the practices currently in place today for glider data management.

Some key considerations for glider data management are summarised below.

- **Data versioning** – for a typical glider deployment we can expect to see multiple versions of the dataset come online, ranging from preliminary datasets whilst the platform is in the water, through to highly refined versions having undergone expert scrutiny post-platform recovery. The resolution of the dataset may also vary between versions, both in terms of sampling frequency and the number of parameters reported. Formalised data management for gliders needs to account for these different dataset streams and implement suitable strategies for superseding and merging.
- **Data standards** - adoption of recognised standards facilitates the discoverability and exchange of glider observations and encourages greater interoperability. Use of controlled vocabularies to annotate metadata and data, together with adoption of recognised data exchange formats must underpin any glider data management strategy.
- **Quality Assurance (QA) and Control (QC)** – both key requirements to add integrity and gain user confidence in datasets. Flexible, consistent approaches that are tailored to specific end user requirements are strongly desirable. QA and QC approaches must be defined, harmonised and documented in a robust infrastructure for glider data management.



- **Data delivery** – timely, efficient data flow that meets all end user requirements is imperative. As with data standards there are recognised data delivery protocols and tools enabling appropriate glider data dissemination to end users.
- **Operational monitoring** – an important component of any data management system incorporating near real time data is the capture of metrics to monitor data availability-accessibility and the health of the data management workflow between stakeholders.
- **Collaborative working** - helps to pool knowledge and develop expertise. It also increases visibility, encourages commonality and reduces duplication of effort. Ultimately it is an important pre-requisite to enable the building of shared infrastructures to pool and disseminate data to stakeholders. Governance is also a key component of collaborative working.

Adoption of best practices enables alignment with FAIR (Findable-Accessible-Interoperable-Reusable) data management principles, one of the Strategic Objectives of GOOS. These topics feature heavily in recent OceanObs'19 white papers such as best practices (Pearlman *et al.*, 2019), Ocean FAIR data (Tanhua *et al.*, 2019), data interoperability (Buck *et al.*, 2019; Snowden *et al.*, 2019) and the future of ocean gliders (Testor *et al.*, 2019). The final recommendations of the OceanObs'19 conference, held in September 2019, underpins the conclusions reached by this deliverable.

1.3. Goals of deliverable and structure for remainder of report

The key objective of this deliverable is to enable glider data assimilation into the SeaDataNet infrastructure. As such, there will be a strong European emphasis to the remainder of the report, albeit with the global context also set, where required. In order to manage glider data effectively within the SeaDataNet infrastructure it is essential to develop expertise in glider data itself and the existing protocols already in place to appreciate the wider picture.

The following chapters will i) provide readers with a short timeline of key events that have helped to develop and fine tune the current glider data management infrastructure, ii) elaborate on the best practice outlined in section 1.2, iii) detail how glider data should be assimilated into the SeaDataNet infrastructure and make recommendations to facilitate the process.

2. A brief history of glider data management

2.1. Europe: EGO to GROOM to OceanGliders to GOOS

The **EGO initiative** was launched in 2005 by a group of oceanographers keen to establish a strong community of European glider users. The main motivation behind this being that a network of experts working in collaboration would be more effective at resolving technical and scientific challenges presented by these relatively new platforms. Creation of a glider community within Europe was also seen as a pre-requisite for the development of a support infrastructure necessary to streamline the flow of glider observations to potential end users. Regular community meetings and associated glider training schools ensured early success for the EGO concept.

An **OceanObs '09 community white paper** (Testor *et al.*, 2010) recognised the value of assimilating glider observations into GOOS. Further momentum to the EGO concept and GOOS aspirations were given a major boost by an **EU COST Action (ES0904)** in 2010, helping to develop greater coordination



sdn-userdesk@seadatanet.org – www.seadatanet.org

SeaDataCloud - Further developing the pan-European infrastructure for marine and ocean data management

Grant Agreement Number: 730960

of glider activity within Europe. The '09 paper has been further developed in a recent OceanObs'19 white paper (Testor et al., 2019).

In 2011 the EU project **GROOM** (Gliders for Research, Ocean Observation and Management, 2011-2014), within the framework of the FP7 'Infrastructures – Design Studies' Call was funded. GROOM, driven by the now established European glider community, was a significant milestone for European glider data management, ratifying a preliminary common data exchange format (the EGO NetCDF format, based on similar format specifications for Argo and OceanSITES) and identifying fundamental workflows for exchange of glider data within the European landscape.

In 2015 the EuroGOOS Glider Task Team was established in an attempt to refocus European efforts in building common glider infrastructure, diminished somewhat since the end of GROOM.

In 2016 **OceanGliders** was launched as an associated program of GOOS (see Figure 1). A dedicated **OceanGliders Data Management Task Team (OGDMMT)**, operating under the program, is the current authority for international data management best practice for glider data observations.

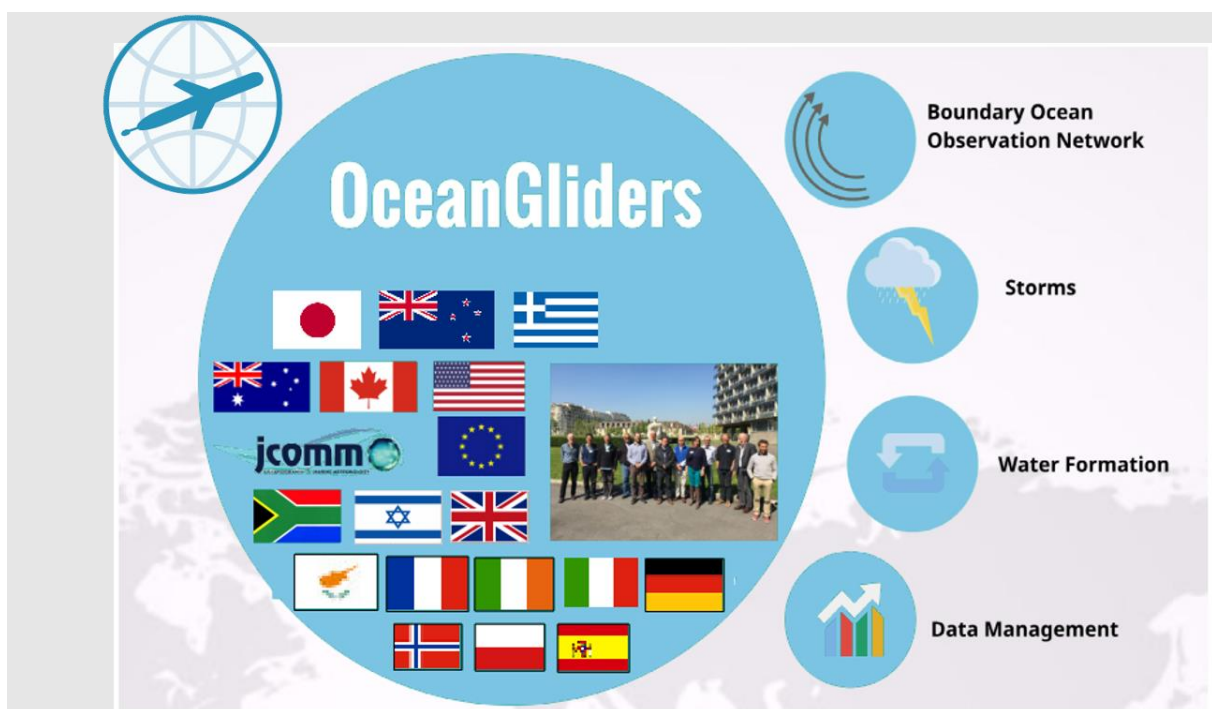


Figure 1: The OceanGliders Program, endorsed by WMO/IOC JCOMM in October 2017. The Program is comprised of four individual Task Teams driving specific domain expertise. One of these Task Teams focuses on Data Management (Source: P. Testor, OceanGliders Chair)

In parallel to EGO, similar glider data management best practices have been developed in the U.S and Australia by IOOS and IMOS respectively. Until fairly recently minimum effort has been channelled into ensuring global interoperability of glider datasets originating from these three main drivers. Fortunately, all three have built their exchange format protocols around the NetCDF data model, which will help future efforts (under the coordination of OGDMMT) to introduce greater global alignment.

2.2. Achieving global interoperability

Under the auspices of the OGDMTT a current priority is to find commonality between the three principal (EGO, IOOS and IMOS) glider data exchange formats allowing end users to bridge between each and aggregate glider data, irrespective of its origins. Global interoperability will play a fundamental role in achieving FAIR compliance. This will be discussed further in the next chapter.

3. Glider data management principles and best practice

This section examines current best practice for glider data management. The focus is on European protocol, but also addresses the global picture, driven largely by OGDMTT, coupled with overall aspirations to achieve FAIR compliance.

3.1. Data versioning

There are different data states that the data user needs to understand. These are detailed below:

3.1.1. Near Real Time (NRT) data stream

These data are telemetered by the glider whilst it is in the water. The data stream may have had real time automated QC applied.

For some glider platforms (e.g. Slocum) the NRT data stream is typically a coarse resolution dataset, comprising a subset of science and engineering variables and at a low sampling frequency.

3.1.2. Recovery (REC) data stream

Data downloaded from the glider upon recovery. The recovery data stream can include a higher resolution version of telemetered data (e.g. Slocum platforms). This data stream can also be accompanied by data from self-logging auxiliary sensors attached to the glider platform (e.g. raw passive acoustic data).

3.1.3. Delayed Mode (DM) data stream

Delayed mode quality controlled data – data fully worked up to a ‘scientific grade’ of quality assurance, this typically is available up to 12 months after a glider is recovered. Delayed Mode data will place additional requirements on the minimum mandatory metadata requirements to ensure quality control is fully documented and self-describing in the data files.

The terminology for data versions is analogous to Argo and OceanSITES and is broadly understood and commonly used within the marine science and data communities. However, the addition of the recovery data version is a new concept.

3.2. Data (and metadata) standardisation

Aligning to community standards for storage, markup and exchange of glider observations enables greater interoperability and paves the way to greater FAIR compliance. In this section, protocols for formatting and markup of glider observations are discussed.

3.2.1. Data exchange formats

Consistent, well designed file formats are essential to facilitate standalone, interoperable information exchange. For gliders the challenge is particularly great given the wealth of science, technical and engineering information acquired from glider platforms, coupled with the requirement to store comprehensive metadata alongside those data. The exact formats chosen to disseminate glider data vary by global region, but are typically based on NetCDF, albeit with some variations.

EGO

Glider data aligned with the EGO infrastructure are currently required to meet the V1.2 specification of the EGO NetCDF file format, which was ratified in May 2017.

EGO V1.2 builds upon a preliminary V1.0 release (published under GROOM in October 2012). An interim release, V1.1, which attempted to introduce increased alignment with other global exchange formats, was ultimately abandoned in favour of presenting the (largely) European community with a stable file format target.

EGO data are recorded as trajectories which are based largely on OceanSITES, Argo and ANFOG user manuals, but with additional glider-specific elements. Incorporation of glider phase information allows for the possibility of subsequent subsetting of the trajectories to individual component water column profiles.

The EGO format is a well-defined standard, which aligns to:

- Climate and Forecast (CF) protocol
- The Cooperative Ocean/Atmosphere Research Data Service (COARDS) NetCDF Conventions
- Unidata units (UDUNITS)
- The Attribute Convention for Data Discovery (ACDD)
- ISO8601 (time)
- SeaDataNet conventions

The EGO format can carry a wealth of metadata and stores data stream versioning and quality control information. Openly available EGO resources include:

- The EGO gliders NetCDF format reference [manual](#) version 1.2.
- The EGO gliders data [processing chain](#), version 20180920_004n (enabling users to convert proprietary Slocum, Seaglider and SEAEXPLORER glider data into the EGO common exchange format).
- The NetCDF file [format checker](#) for Argo floats, Copernicus In Situ TAC, EGO gliders, OceanSITES (enabling users to validate output from the EGO processing chain).
- The SOCIB toolbox



IOOS and IMOS

Data exchange formats in the U.S and Australia are governed by IOOS and IMOS (ANFOG) respectively. Both have adopted CF-compliant NetCDF solutions. The IOOS format is designed to facilitate the use of Application Programming Interface (API) endpoints, such as ERDDAP for data dissemination. The IMOS format has been engineered to meet specific data integration requirements within Australia, with data primarily distributed by a THREDDS catalogue.

OG1.0 – global interoperability

The existence of three principal exchange formats presents a barrier to the international sharing of glider data and therefore achieving FAIR. The formation of OGDMTT was seen as an opportunity to harmonise to a single global exchange format. A bespoke OGDMTT Data Format Working Group was established in 2019 to achieve this aim. This original target of a unified common format, termed OG1.0, was considered unrealistic after the preliminary Working Group meeting took place in May 2019. The current solution is to work towards agreeing a minimum set of mandatory metadata and simple set of variables common between the three formats. This would in effect create a metadata and variable best practice guidance to ensure files are interoperable at an international level with data delivery tools and allow global interoperable exchange of key glider observations. Individual glider data files meeting the minimum set of metadata defined by OGDMTT would be labelled 'OG1.0 compatible'.

3.2.2. Controlled vocabularies

This sub-section will focus on the usage of controlled vocabularies in the markup of metadata and data within the EGO data exchange format, used for data dissemination by DACs aligned to the GDAC, Coriolis.

EGO Global Attributes

Although free-text fields are permissible within certain EGO global attribute fields, where possible controlled vocabulary terms are adopted. These include:

- WMO platform codes – unique numeric identifiers assigned to distinguish glider platform instance. Assignment of WMO platform codes are strictly governed. For EGO gliders, JCOMMOPS are the authority able to assign these. WMO code assignment is a pre-requisite for distribution of glider observations within the WMO GTS system.
- ICES platform codes – unique platform instance identifiers administrated by ICES (with approval by the National Oceanic and Atmospheric Administration, NOAA). Content governance of the sub-surface glider platform class (ICES platform class 27) is performed by the British Oceanographic Data Centre (BODC), who are members of OGDMTT. ICES platform codes underpin the SeaDataNet C17 vocabulary with a widespread global user base.
- EDMO – the European Directory of Marine Organisations, one of the SeaDataNet discovery metadata catalogues, currently containing more than 4000 organisations globally.
- EGO Reference Tables – these are controlled vocabularies which borrow heavily from the existing Argo Data Management infrastructure and include (although are not limited to) terminology required to markup autonomous profiling platforms.
- ACDD global attribute names



EGO variables

Attributes of EGO variables also utilise controlled vocabularies. These include:

- CF Standard Names (where available)
- NVS P01 and P06 – the BODC Parameter Usage Vocabulary and BODC-approved data storage units, respectively.
- EGO parameter names, re-using the existing Argo observed [parameter list](#), where possible. To constrain EGO parameter names more robustly the OG1 vocabulary was introduced by OGDMTT (see below).

Ocean Glider Network Parameter Usage Vocabulary (OG1)

This vocabulary was established in order to formalise EGO parameter name terminology and provide a direct mapping to NVS P01 and P06 terms, and CF Standard Names listed in the NVS vocabulary P07. This paves the way for future mappings to existing (and new) GOOS Essential Ocean Variables (EOVs), and UD units, enabling the full potential of glider data assimilation into GOOS to be realised. A **subset** of Essential Ocean Variables and Essential Climate Variables (ECVs), assembled by AtlantOS (and a useful starting reference for gliders), are listed in the NVS vocabulary A05.

NVS OG1 vocabulary content is governed by OGDMTT. Figure 2 contains an example of an entry in the OG1 vocabulary – in this case mapping the EGO ocean glider observed property ‘PSAL’ to P01 term ‘PSALST01’/P06 term ‘UUUU’.

↑ -- Practical salinity --

| | |
|------------------------|---|
| URI | http://vocab.nerc.ac.uk/collection/OG1/current/PSAL/ |
| Identifier () | SDN:OG1::PSAL |
| Preferred label (en) | Practical salinity |
| Alternative label (en) | PSAL |
| Definition (en) | Practical salinity of the water body by CTD and computation using UNESCO 1983 algorithm |
| Version Info () | 1 |
| Has Current Version | http://vocab.nerc.ac.uk/collection/OG1/current/PSAL/1/ |
| PAV Version () | 1 |
| PAV Authored On () | 2018-05-31 16:35:24.0 |
| Deprecated() | false |
| Narrower | http://vocab.nerc.ac.uk/collection/P01/current/PSALST01/ |
| Related | http://vocab.nerc.ac.uk/collection/P06/current/UUUU/ |
| Date () | 2018-05-31 16:35:24.0 |

Figure 2: Example of OG1 vocabulary entry

3.3. Data Quality Control (QC) and Quality Assurance (QA)

Robust strategies for applying and documenting QC and QA are important to ensure glider observations are trusted by data assimilators and users. Unfortunately, at present, approaches to QC and QA vary by region and by DAC.

As with Argo, there are two main tiers of QC agreed by the glider community:



3.3.1. Real Time Quality Control (RTQC)

These are based largely on the established QC routines developed and applied by the Argo community. Several groups have tailored these to meet specific glider community requirements, including Ifremer, SOCIB, IOOS and ANFOG. Some of these groups have made their tools publically available (see Table 2 for further details).

3.3.2. Delayed Mode Quality Control (DMQC)

A much more complex subject, which has engaged (and occasionally divided) scientists and data managers for many years. DMQC is expected to deliver high quality datasets. Although efforts have been made to define DMQC standards for glider data, this is still considered in its infancy. Examples of published methods for the DMQC of glider data are Cummings et al (2011) and Garau (2011).

A similar exercise, carried out by Argo, took some time to establish DMQC protocols and for these to gain widespread acceptance. It is anticipated that the process may be lengthy for gliders too, but efforts should be made to learn from the Argo example.

3.3.3. Harmonisation

Whilst EGO have published **protocols** for handling glider QC/QA, OGDMTT recognises the need to revisit these and define a clearer global strategy, which aggregates best practice from across the global community. With this goal in mind OGDMTT recently established a Working Group, with global membership, to define OceanGliders quality procedures.

3.4. Data exchange pathways and tools

This section describes the principal exchanges of glider information currently taking place between glider stakeholders and the technologies underpinning these.

3.4.1. The Global Data Assembly Centre (GDAC) model

The GDAC model forms the basis of glider data management in Europe, facilitating exchange of all three data streams (NRT, REC and DM) to stakeholders. The information flow is illustrated in a schematic prepared by OGDMTT - Figure 3 below. Glider data and metadata (including deployment discovery metadata) are fed by glider operators and Principal Investigators (PIs) to designated Data Assembly Centres (DACs). In many cases the DAC is also a National Oceanographic Data Centre (NODC). Exchange of information in common exchange formats between PI and DAC is encouraged. Individual DACs collate these glider datasets (and any subsequent versions) and push these to the EGO glider GDAC, based at Coriolis, Ifremer. The GDAC helps to provide a centralised point from which to serve glider observations. In Europe, the GDAC is a primary conduit for glider observations to operational stakeholders within the European operational oceanographic landscape.



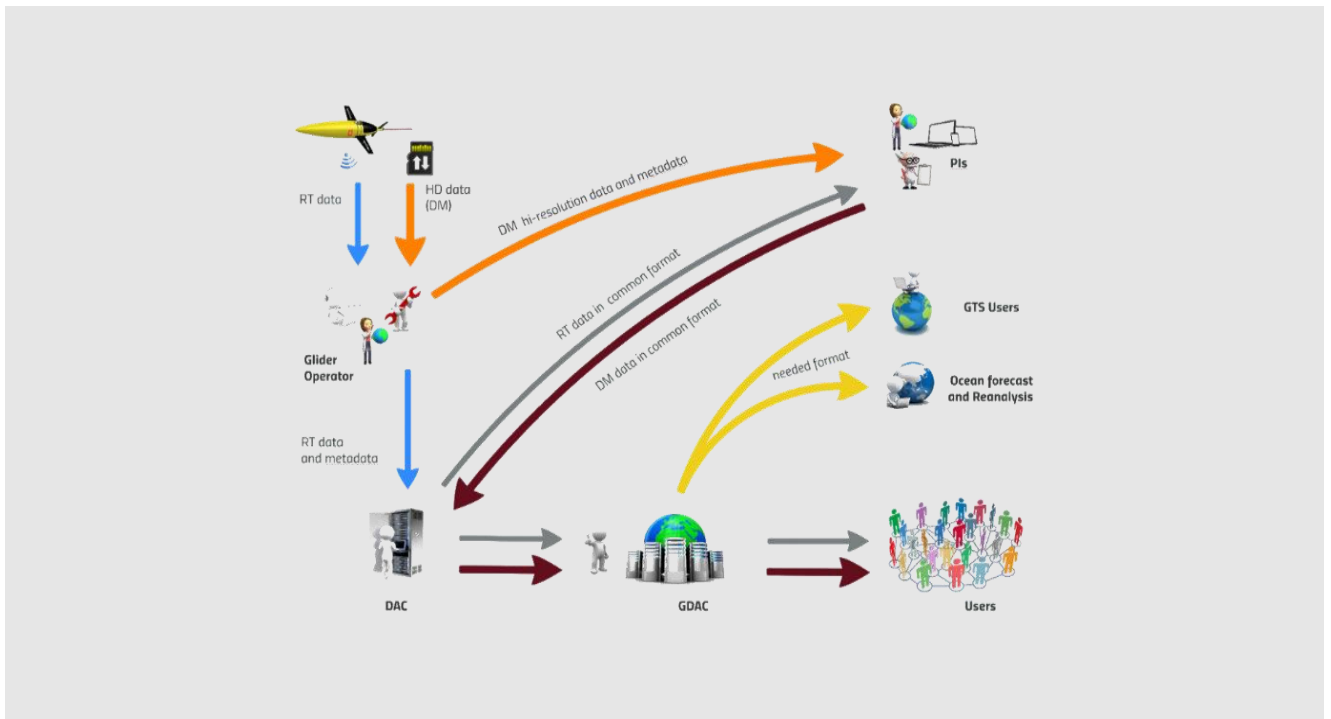


Figure 3: Data flow of glider data within the DAC-GDAC framework, as identified by OGDMTT. This is an idealised model. Some variations exist, such as some DACs coordinating the GTS push themselves.

Outside of Europe there are two counterparts to the EGO GDAC, IOOS in the U.S and IMOS in Australia. Although IOOS and IMOS are considered large DACS, not GDACS, collectively all three play an influential part in setting and steering international protocols for glider data exchange. Key information about all three organisations are summarised in Table 1 below.

Table 1: Operational glider GDAC/DACs, exchange formats and key data delivery protocols

| Organisation | Hosting institution | Data exchange format | Data delivery mechanisms |
|---------------------|-----------------------------------|----------------------|--|
| EGO GDAC | Ifremer, France | EGO NetCDF | EGO FTP area |
| IOOS US glider DAC | NOAA, USA | IOOS NetCDF | ERDDAP, IOOS data portal, NOAA one stop web portal |
| IMOS Australian DAC | University of Tasmania, Australia | ANFOG/IMOS NetCDF | THREDDS, AODN data portal |

Harvest and integration of (active) glider deployment metadata from the three assembly centre authorities has been performed by OGDMTT to aid discoverability of global glider activity and help achieve FAIR-compliance – see Figure 4 below.



Figure 4: Example of active global gliders (July 2019) discoverable via the OceanGliders website.

3.4.2. GTS (Global Telecommunication System) data assimilation

Timely (< 24 hours) NRT data exchange with the GTS is an integral part of the global glider data management framework. This is typically achieved by GDAC or DAC push of standardised data to approved operational ocean forecasting agencies, who, in turn, make data available to the GTS.

At present, profiles of open-access glider data are assimilated into the GTS in WMO TESAC or BUOY format. The aim is to move towards the more comprehensive WMO BUFR format, which will handle finer (spatial/temporal) resolution data for a wider range of variables. Under OGDMTT a dedicated Working Group was formed in 2019 to produce a glider-BUFR template, in collaboration with operational stakeholders.

GTS-assimilated data ultimately feed into the Global Temperature and Salinity Profile Programme (GTSP) repository for long term archive.

3.4.3. THREDDS and ERDDAP

THREDDS, developed by Unidata (part of the University Corporation for Atmospheric Research, UCAR) facilitates retrieval of remote, structured data and metadata via web services. THREDDS is underpinned by various remote data access protocols. IMOS in Australia maintain a THREDDS catalogue as a fundamental component of their data dissemination strategy.

ERDDAP is also a web-based data server promoting easy access to and sharing of consistent, unified scientific data. ERDDAP has been developed by the U.S. National Oceanic and Atmospheric Administration (NOAA) and underpins glider data delivery within the U.S. It is likely that ERDDAP will play an increasingly central role in global glider data interoperability and achieving FAIR compliance (Snowden et al., 2019).

3.4.4. Open Geospatial Consortium (OGC) Standards and Sensor Web Enablement (SWE)

Open Geospatial Consortium (OGC) Sensor Model Language (SensorML) and World Wide Web Consortium (W3C) Semantic Sensor Network (SSN) Ontology are standardised machine-readable metadata standards. They potentially allow automated sharing of sensor metadata and data between sensors and processing systems, sensor networks and research infrastructures. With the sensors installed on gliders having the potential to be used on other platforms and data integration needed between glider data and that of other networks, such as Argo, a common metadata standard would facilitate the sharing of such metadata.

As part of the European Union FP7 ‘The Ocean of Tomorrow projects’ (2010-13) and Ocean Data Interoperability Platform (ODIP) projects (2012-18), a common Marine Sensor Web Enablement (SWE) profile for SensorML was developed.

Adoption of SensorML is ongoing within the EMODnet Data Ingestion project for the (near) real time acquisition of data from providers and the ENVRIplus project Deliverable 9.1 (TC_4 sensor registry) demonstrates the potential for cross research infrastructure sharing of sensor metadata using SensorML.

Adoption of SSN to data has been limited in the marine community but this potentially offers interoperability of metadata in the internet of things concepts along with the ‘new’ OGC SensorThings API which has not been trialled on marine data to the authors of this deliverable’s knowledge.

3.5. Common community tools for processing and quality control of glider data

Within the glider community a number of groups have made their tools for the processing and delivery of glider data openly available. A summary of tools that the authors of this document are aware of is shown in Table 2 below (including some discussed previously in this section, but included again for completeness). The move towards open software and electronic notebooks such as Jupyter in the research community means that there are a multitude of tools with varying capability being made freely available, as such we recommend a comprehensive list would need to be maintained by the OceanGliders program on behalf of the community.

Table 2: Table summarising shared glider processing, QC and delivery tool

| Tool | Data processing | Data QC | Data Delivery to GDAC | Link to tool |
|--------------------------------------|-----------------|---------|-----------------------|---|
| SOCIB, glider toolbox | ✓ | ✓ | ✓ | https://github.com/socib/glider_toolbox |
| Ifremer, EGO glider processing chain | ✓ | ✓ | ✓ | https://www.seanoe.org/data/00343/45402/ |
| UEA toolbox | ✓ | ✓ | ✗ | http://www.byqueste.com/toolbox.html |
| Gliderscope software | ✓ | ✓ | ✗ | http://imos.org.au/facilities/oceangliders/glider-data/gliderscope/ |



| | | | | |
|-----------------------------|---|---|---|---|
| IOOS glider DAC software | ✓ | ✓ | ✗ | https://github.com/ioos/ioosngdac |
|-----------------------------|---|---|---|---|

4. Integration with SeaDataNet

4.1. Rationale for integration

A formal link between the EGO and SeaDataNet infrastructures is, in many ways, the missing piece of the jigsaw for glider data management at the European scale. There are likely to be many benefits:

- Expansion of the EGO community as SeaDataNet partners align and mobilise their observations. This would further consolidate the EGO initiative – a core element of the global OceanGliders program.
- A pipeline of archival (i.e. Delayed Mode) datasets to SeaDataNet adding to the richness of data holdings (and data products). Not only more datasets, but datasets which are often from regions and seasons which are under-sampled by traditional sampling methodologies.
- Integration into large European e-infrastructures such as SeaDataNet (and consequently EMOdnet) enables a broader uptake of data beyond the ocean glider community, increasing the FAIRness of glider data and improving the flow of observations within the wider European marine data management network.

4.2. AtlantOS recommendations for integration

Work Package 7 of EU Horizon 2020 project AtlantOS (2015-2019) considered how to improve the European data flow and data integration from a variety of data networks. As part of a dedicated community workshop in December 2016, the mechanics of Delayed Mode glider data into SeaDataNet, as a data integrator, were considered in a community [report](#). The report concludes that glider data assimilation should follow a very similar model to that proposed for Argo, which encompassed: i) creation of Common Data Index (CDI) metadata files from Argo data, ii) full adoption of EDMO to facilitate discoverability, iii) creation of Ocean Data View (ODV) files to serve Argo data, iv) introduction of mechanism to avoid Argo data duplication within the SeaDataNet infrastructure.

4.2.1. Implementation of AtlantOS recommendations for Argo

SeaDataCloud has already commenced integration of Argo datasets into the SeaDataNet infrastructure, as part of WP9. The default, as of 01 January 2019, is for Ifremer to pull Argo metadata and data into SeaDataNet on behalf of SeaDataNet partners directly from the central Argo GDAC database (also based at Ifremer). This approach was considered to be more effective and less prone to inconsistency than individual SeaDataNet/Argo DAC partners overseeing the migration.

A mapping between the Argo format and SeaDataNet CDI profile was performed and is the principal mechanism for Argo dataset discovery within SeaDataNet. Expansion of EDMO to support Argo data delivery, via SeaDataNet, was performed by JCOMMOPS.



sdn-userdesk@seadatanet.org – www.seadatanet.org

SeaDataCloud - Further developing the pan-European infrastructure for marine and ocean data management

Grant Agreement Number: 730960

The metrics gathered for Argo data availability within SeaDataNet are based on dataset custodians (i.e. the individual SeaDataNet partners/Argo DACs) and not Ifremer to fairly attribute correct provenance of the data.

Argo data are distributed in SeaDataNet NetCDF and ODV format by SeaDataNet, with the conversion from the Argo file format being handled internally at Ifremer.

Figure 5 below shows an amended version of Figure 3, augmented to incorporate some key European stakeholders (source OGDMTT). CDI is the recommended mechanism for exposing Delayed Mode Argo and glider data within SeaDataNet to data aggregators, such as EMODnet Physics.

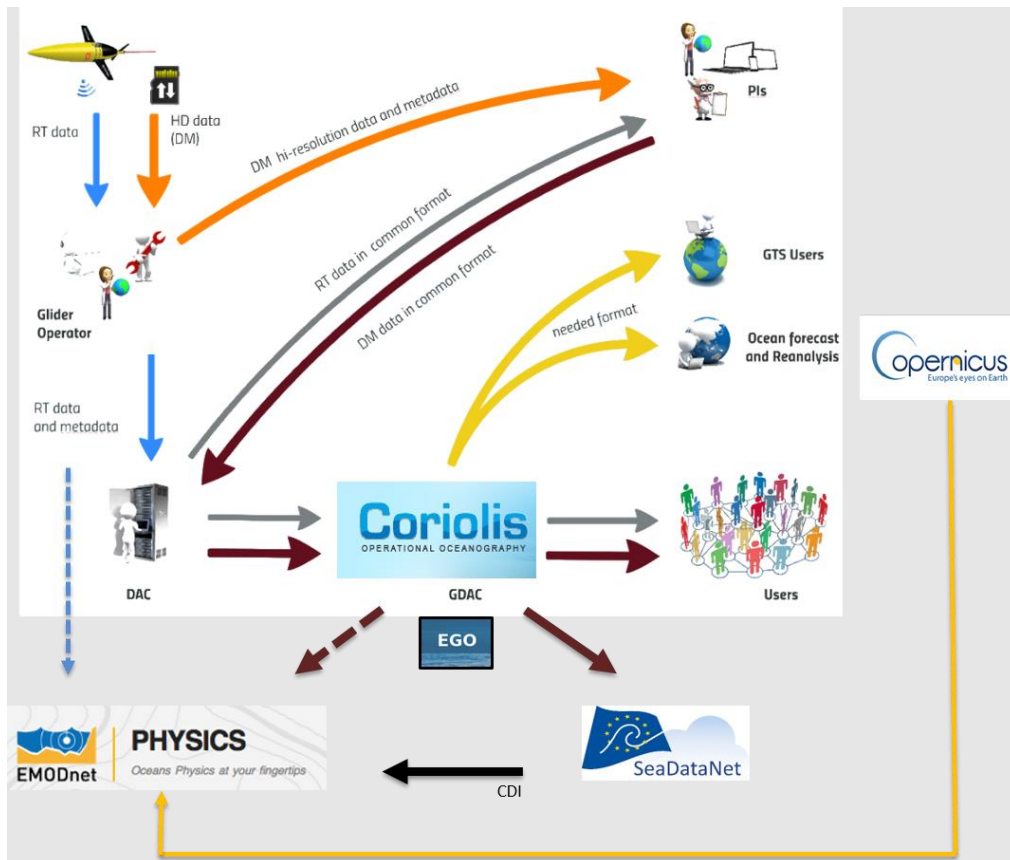


Figure 5: SeaDataNet incorporated into an idealised flow of data observations (including Argo and gliders) within Europe.

For gliders we have the added complexity of assimilating restricted datasets, which are held at DACs, but not pushed to the EGO GDAC. The recommendation here would be for SeaDataNet partners/DACs to assume responsibility for preparing and submitting CDIs (if metadata are derestricted), thereby bypassing the GDAC pipeline. It is unlikely that the GDAC will handle restricted datasets given that it is very much aligned to the open-access requirements of GOOS.

4.3. Mapping the EGO data exchange format (V1.2) to CDI

Following the recommendations from AtlantOS, a fundamental exercise required to bring glider data into SeaDataNet is population of the CDI standard. Table 3 below shows how the CDI profile could, in part, be fulfilled from information already packaged up within the EGO V1.2 data exchange format. The table also captures recommendations to further align CDI and EGO. Mandatory fields in CDI and EGO are in bold characters.

Table 3: Proposed mapping between SeaDataNet CDI profile and EGO format V1.2

| CDI field | Mapped to EGO attribute | Comments/Recommendations |
|---|-------------------------|---|
| cdi-identifier | id | id = EGO filename, without .nc suffix |
| ISO 19139 header xml header | - | default |
| METADATA CREATING ORGANISATION | - | Populate with: GDAC (Coriolis), as default |
| METADATA CREATION-DATE | - | Populate with: CDI creation date |
| Metadata Standard Name | - | default |
| Metadata Standard Version | - | default |
| MEASURING AREA TYPE | - | Populate with: surface |
| SPATIAL REPRESENTATION HORIZONTAL RESOLUTION VERTICAL RESOLUTION TIME RESOLUTION | - | Not specified in EGO (same for Argo approach) |
| DATUM OF COORDINATE SYSTEM | - | Populate with: WGS84 - EPSG::4326 |
| Metadata Extension info | - | default |
| NAME/ALTERNATIVE NAME OF THE DATASET | Id | id= EGO filename, without .nc suffix |
| DATASET-ID | Id | id = EGO filename, without .nc suffix |
| REVISION-DATE OF DATASET | date_update | - |
| IDENTIFIER | Id | id = EGO filename, without .nc suffix |
| ORIGINATORS OF THE DATASET | sdn_edmo_code | - |
| ABSTRACT ON DATASET | title summary | Use summary from 'title summary' field – in line with typical CDI abstracts |

| | | |
|--|-------------------------------------|--|
| ORGANISATION MANAGING THE DATASET | data_assembly_center | Mapping needed between EGO reference table 4 and EDMO, but SDN recommendation should be for EGO to fully adopt EDMO |
| RESOURCE MAINTENANCE | update_interval | - |
| INSPIRE reference | - | default |
| PARAMETERS | - | Harvest P02 from SensorML (where available) or from P01 in data file |
| INSTRUMENT and POSITIONING SYSTEM | - | Harvest L22 from SensorML (where available) |
| PLATFORM | - | Harvest L06/B76 from SensorML (where available) or populate from L06 (e.g. 'autonomous underwater vehicle' for profiling gliders) May need to map JCOMMOPS platforms codes to NVS to allow a more accurate platform description |
| PROJECTS | PROJECT_NAME | SDN recommendation for EGO to adopt European Directory of Marine Environmental Research Projects (EDMERP) |
| Use Limitation | distribution_statement and citation | - |
| DATASET ACCESS RESTRICTIONS | - | Populate GDAC-sourced datasets with: UN (unrestricted – all EGO published data are open access). Populate DAC-sourced restricted datasets with: RS (by negotiation) |
| STATION NAME and/or CRUISE NAME | DEPLOYMENT_CRUISE_ID | - |
| EDMED REFERENCE | - | Recommendation to extend EGO format with an optional field to further align with SDN |
| CSR REFERENCE | DEPLOYMENT_CRUISE_ID | Populate with Cruise Summary Reports (CSR) CSRREF corresponding to DEPLOYMENT_CRUISE_ID |

| | | |
|---|--|--|
| | | Note: many gliders are deployed from small boats, fishing vessels etc., so this may need to be optional |
| SPATIAL RESOLUTION | - | Not specified in EGO |
| Language used with the dataset | - | default |
| Character set | - | default |
| Main theme of the dataset | - | default |
| GEOGRAPHICAL COVERAGE WEST | geospatial_lon_min | - |
| EAST | geospatial_lon_max | - |
| SOUTH | geospatial_lat_min | - |
| NORTH | geospatial_lat_max | - |
| TRACKS (Curves) | - | - |
| AREAS (Surfaces) | - | Populate with LONGITUDE_GPS/LATITUDE_GPS pairs |
| START AND END DATE (AND TIME) | time_coverage_start/time_coverage_end | - |
| MINIMUM DEPTH OF OBSERVATION MAXIMUM DEPTH OF OBSERVATION WATER DEPTH VERTICAL DATUM | geospatial_vertical_min geospatial_vertical_max | Water depth – not specified in EGO (or Argo, which use -9999) Populate VERTICAL_DATUM with: sea level (same as Argo approach) |
| ADDITIONAL DOCUMENTATION (PUBLICATION) | - | Not specified in EGO |
| ORGANISATION DISTRIBUTING THE DATASET | - | Populate with data_assembly_center to ensure metrics report at the DAC and not the GDAC level. |
| Dataformat Version | - | Populate with: CFPOINT |
| DISTRIBUTION INFO/SERVICE BINDINGS Data size | references | Default and EGO attribute 'references', where available |

| | | |
|---|-----------|-----------------------|
| Distribution website Distribution protocol Database reference Distribution Method | | |
| Data Quality Information Scope Report – Name Report – Date Report – Comment Report – Status Lineage | qc_manual | Same as Argo approach |

4.3.1. Key observations from CDI-EGO mapping

A complete mapping between CDI mandatory fields and EGO V1.2 is possible, but many of the required fields are not currently mandatory in EGO, which could pose a challenge for SeaDataNet integration. There are also clear opportunities for EGO to align more closely with SeaDataNet discovery services (**EDMED**, **EDMO**, **CSR**, **EDIOS** and **EDMERP**). This will be a key recommendation to take to the OGDMTT, more specifically the Working Group on data formats. The exercise to identify a minimum set of metadata between EGO, IOOS and IMOS (i.e. ‘OG1.0’) presents an opportunity to push the SeaDataNet agenda too.

EGO V1.2 recommends EGO global attribute ‘deployment_code’ is utilised to fulfil the function of CDI local identifier for SeaDataNet, but the recommendation here is to adopt global attribute ‘id’ instead, which currently presents a more unique file identifier.

Widespread adoption of OGC-SWE Standards, such as SensorML, would present an opportunity to auto-harvest information desirable for CDI. Indeed, the creation and maintenance of a SensorML catalogue for gliders was a recommendation by AtlantOS. The template to achieve this was developed as part of ‘The Ocean of Tomorrow projects’ and is freely available in a GitHub [repository](#). This has been implemented by SeaDataCloud BODC and has platform and sensor metadata available in SensorML. The template has also been applied in the freely available **52North** Sensor Observation Service (SOS) server software.

It should be noted that the EGO format is in a period of flux and CDI mapping will need to keep afoot of this. New versions of the EGO data exchange format (i.e. OG1.0) are anticipated in the coming months. SeaDataCloud partners on OGDMTT will be an important mechanism for reporting back.

4.4. Near real time pathways of glider observations into SeaDataNet and related European infrastructures

Besides the traditional integration of archive (Delayed Mode) datasets into SeaDataNet, via the CDI, there are other routes for integrating glider observations into our infrastructure, which will benefit SeaDataNet, data providers and the wider sphere of European marine data managers and users:

4.4.1. OGC-SWE route: The 'shop window' approach

SeaDataCloud Deliverable D9.9 - 'Specifications of the SWE Ingestion Service, including SWE profiles and architecture' - has already outlined how the OGC-SWE principles discussed earlier (section 3.4.4) can be harnessed by SeaDataNet to connect directly with oceanographic sensors transmitting observations (see Figure 6 for a summary workflow schematic). The possibility of auto-harvesting basic metadata for CDI population is also discussed in this deliverable. Machine-to-machine transfer of (near) real time information in this fashion can potentially mobilise new data providers within the European marine data infrastructure.

In parallel EMODnet Physics and EMODnet Data Ingestion have been promoting OGC-SWE pathways with the launch of a pilot OGC-SWE **demonstrator** for (near) real time data **exchange**. SeaDataNet has a key role to play in further developing this pilot scheme and ensuring expansion (underpinned with appropriate standards) with additional operators of (near) real time monitoring networks, including those involving gliders.

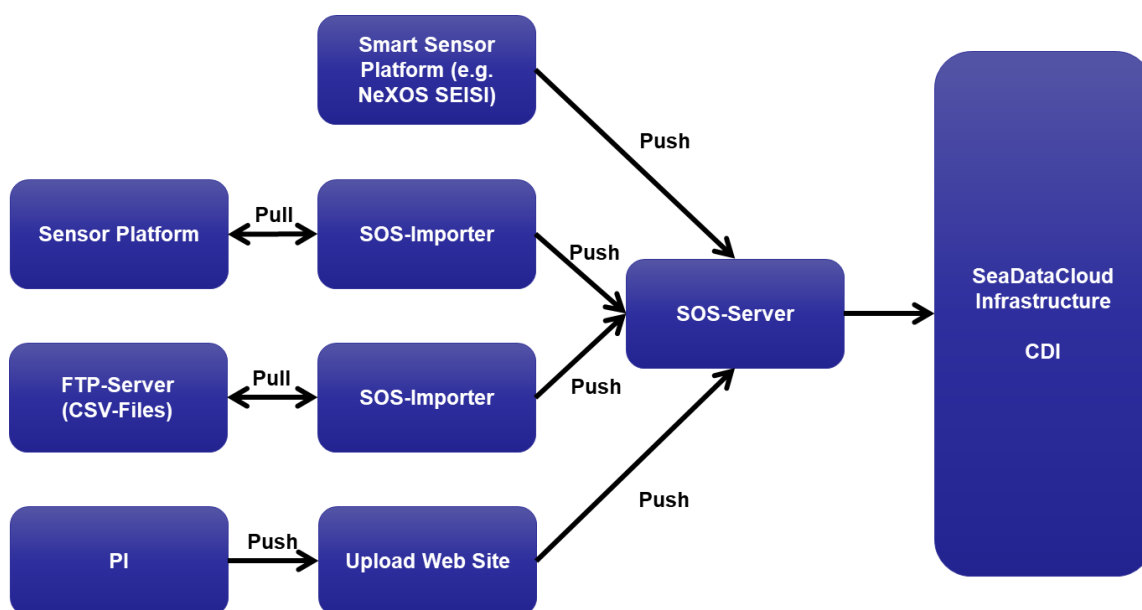


Figure 6: Schematic of OGC-SWE workflow linking data provider directly to SeaDataNet 'shop window' in (near) real time

4.4.2. Assimilation of NRT data collections

As part of SeaDataCloud WP9.6 there is the potential to assimilate further near real time glider observations into SeaDataNet at the data collection level. This activity, aligned with the SeaDataCloud Memorandum of Understanding (MoU) between SeaDataNet and CMEMS, will help to enrich European data products and promote adoption of robust data standards within the overall European marine data infrastructure. The EMODnet Data Ingestion portal and SeaDataNet brokerage service will be important tools to achieve this goal. The latter potentially also enabling assimilation of glider observations from IOOS and IMOS.

The various pathways for exchange of glider observations discussed in this section can be summarised in the workflow diagram presented in Figure 7 below – a schematic underpinning SeaDataCloud WP9.

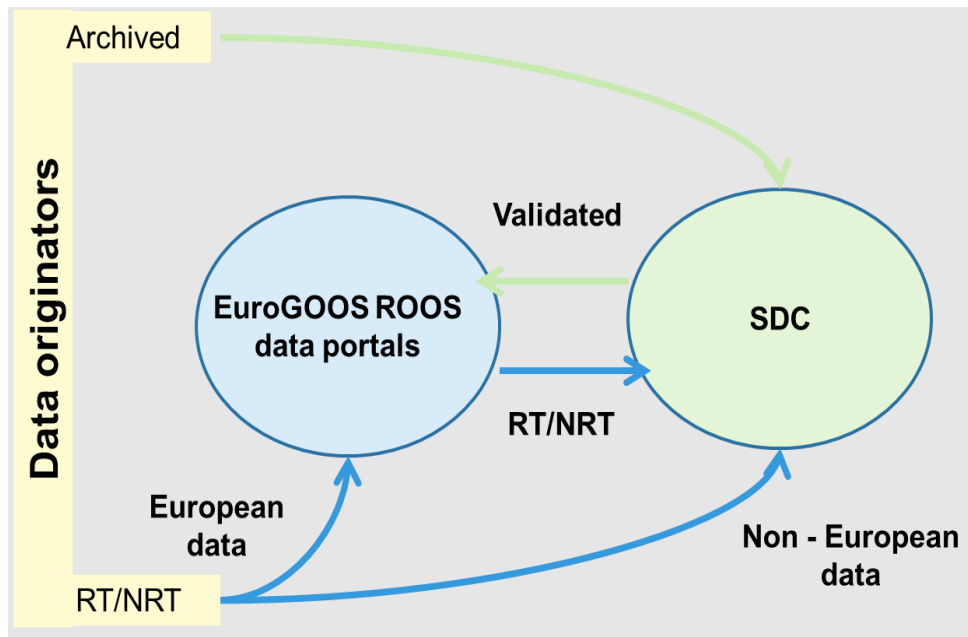


Figure 7: Pathways of data flow to and from SeaDataNet as envisaged by SeaDataCloud WP9

4.5. Data dissemination by SeaDataNet

Following the recommendations of AtlantOS, the population of the CDI catalogue for gliders should promote discoverability of SeaDataNet glider data holdings. The recommendation is to further adopt the SeaDataNet data delivery formats (SeaDataNet NetCDF and ODV) for serving glider data, in a similar fashion to Argo. For open access glider data it is envisaged that transformation from EGO to SeaDataNet NetCDF and ODV will be handled internally by Ifremer, as the conduit between EGO GDAC and SeaDataNet. The assumption is made that this exercise will be readily achievable, following the Argo example (the Argo data model aligning very closely with the EGO data model). Given the transient nature of EGO V1.2, this exercise should be carried out in close collaboration with OGDMTT, in particular their Working Group on data formats.

An additional workflow could be implemented to assimilate restricted glider datasets within SeaDataNet (see section 4.2.1). The suggestion is for SeaDataNet partners/DACs to assume responsibility for submitting CDIs to SeaDataNet for any restricted datasets they handle, assuming the metadata can be freely disseminated.

Formal integration of SeaDataNet within the global glider data management network is illustrated in Figure 8 below and represents a significant advancement in the overall infrastructure.

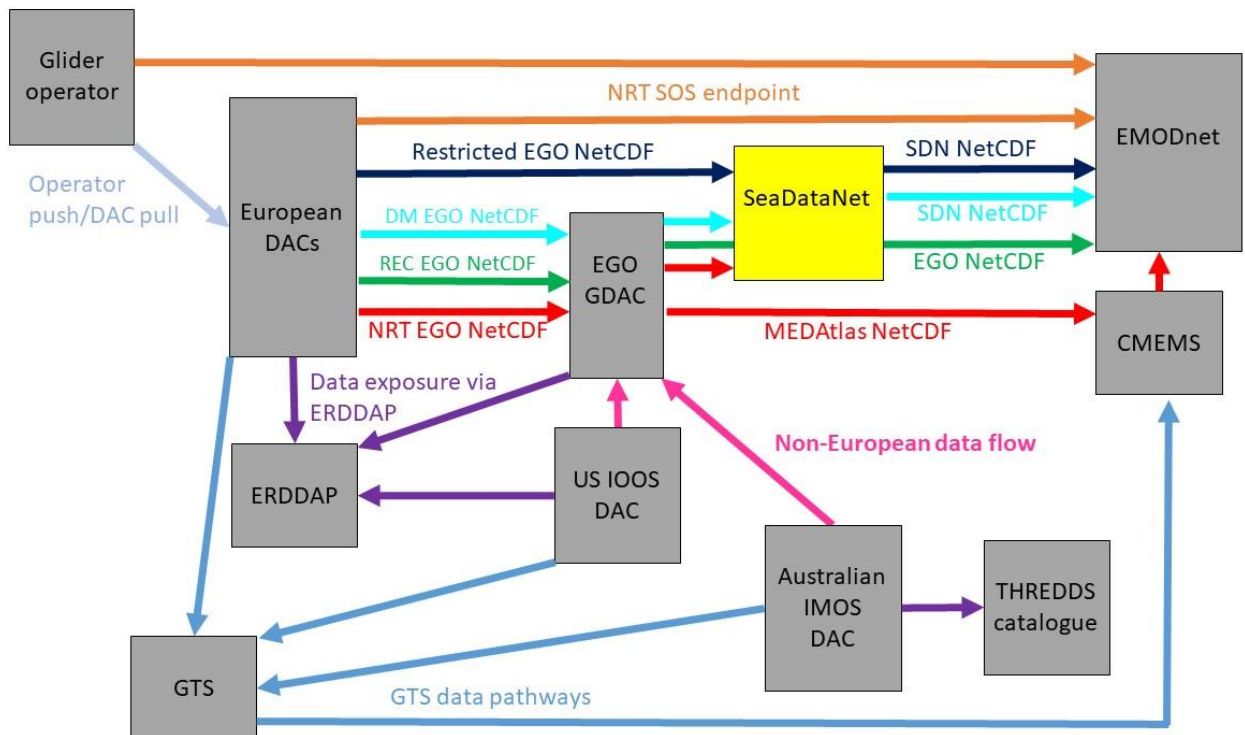


Figure 8: Proposed global glider data pathways with SeaDataNet integration

4.6. Summary of recommendations and concluding remarks

The fundamental recommendation of this report is for SeaDataNet to **align with the EGO initiative**, which has been carefully designed to enable effective exchange of glider data and metadata, particularly within Europe.

The desire to establish glider data management best practice and ensure global FAIRness of glider data has never been so strong. The glider community now has a vehicle, in OGDMTT, to enable this to come to fruition. The work currently being undertaken by OGDMTT, particularly in its format and QC Working Groups, is highly relevant to SeaDataNet and needs monitoring closely. A key recommendation is for **SeaDataNet partners formally connected to OGDMTT to provide an appropriate voice for SeaDataNet requirements** (and a conduit back to SeaDataNet). Current Working Group activity presents an opportunity to **align mandatory global metadata requirements to the needs of SeaDataNet**, in particular the population of the CDI standard. As a minimum, the European EGO standard should be steered towards greater adoption of i) SeaDataNet discovery services and ii) NVS SWE controlled vocabularies, to augment P01 and P06, already underpinning controlled vocabulary markup of the data exchange format. **SWE alignment** will help considerably with the assimilation of sensor information essential in the mobilisation of near real time glider observations.

To gather momentum on the European scale the recommendation is to **pursue pilot demonstrators** – expanding the existing scheme implemented under EMODnet Data Ingestion and EMODnet Physics to push the OGC-SWE agenda. Also a **pilot demonstrating the flow from the EGO infrastructure into SeaDataNet**, via the CDI route advised in this report, would be useful.

Regarding data dissemination, the recommendation is to **mirror the Argo SeaDataNet approach** and serve data in SeaDataNet NetCDF and ODV formats. The **implementation of ERDDAP** by SeaDataNet is worth considering also, as this is a data delivery tool already underpinning glider data management within IOOS and is likely to play an increasingly pivotal role in data dissemination between the three principal global glider authorities (i.e. EGO, IOOS and IMOS)

This report has focused on describing pathways for open access data (a key principal underlying the EGO infrastructure). In order to **aggregate restricted glider data** into data products there would likely be the need to bypass traditional EGO pathways, which are very much aligned to GOOS open data requirements. The logical solution here would be for SeaDataNet glider DACs to directly connect with SeaDataNet, via the CDI route, following agreement from data providers.

An important aspect of glider data management best practice is **persistent digital identifiers**, but this will not be addressed in this report as there are wider discussions on this subject taking place within OGDMTT.

At the time of writing this report international protocols for glider data management are under detailed community review and are likely to change considerably. Some of the recommendations to SeaDataCloud will need to be reassessed when greater stability is reached. It is anticipated that agreement on the minimum metadata underpinning OG1.0 and updates to regional exchange formats (such as EGO) will be completed by the end of 2019.

This review has aimed to provide greater clarity on the fundamental pathways for glider data flow within Europe and beyond. There are, however, a number of **open questions** which need to be addressed before optimal integration can be achieved. These questions, some of which will require consultation with the wider glider community, are summarised below:

- Identified gaps in the CDI mapping to EGO file formats and how to fill these
- The CDI mapping and associated EGO-SeaDataNet file interoperability is an assessment in this deliverable to be fully developed, if the recommendations are adopted
- Clarification of the roles of different European stakeholders
- To ensure data pathways and products are clearer to data users
- Establish efficient data community linkages, brokering and data transfer; to minimise duplication and complexity (this could possibly include the formalisation of a glider SWE.MarineProfile)
- Possible requirement to define NRT-CDI, as an evolution to SensorML/MarineProfile
- The deliverable has not covered strategies for high volume datasets or self-logging data collected by gliders such as passive acoustic data or micro-structure data. There are no current plans to incorporate these in EGO file formats so they may need to be managed separately.
- As already highlighted in section 1.2 the OceanObs '19 conference should complement this deliverable with its theme of information, interoperability, innovation and integration. The findings of this report are rightly biased towards European data flows and the OceanObs'19 activity should help to clearly define data pathways to share and discover data beyond Europe increasing the impact and utility of ocean glider data globally.

5. References

- Buck JJH, Bainbridge SJ, Burger EF, Kraberg AC, Casari M, Casey KS, Darroch L, Rio JD, Metfies K, Delory E, Fischer PF, Gardner T, Heffernan R, Jirka S, Kokkinaki A, Loebel M, Buttigieg PL, Pearlman JS and Schewe I (2019) Ocean Data Product Integration Through Innovation-The Next Level of Data Interoperability. *Front. Mar. Sci.* 6:32. doi: 10.3389/fmars.2019.00032
- Cummings J.A. (2011) Ocean Data Quality Control. In: Schiller A., Brassington G. (eds) *Operational Oceanography in the 21st Century*. Springer, Dordrecht
- Garau, B., S. Ruiz, W.G. Zhang, A. Pascual, E. Heslop, J. Kerfoot, and J. Tintoré, 2011: **Thermal Lag Correction on Slocum CTD Glider Data**. *J. Atmos. Oceanic Technol.*, 28, 1065–1071, <https://doi.org/10.1175/JTECH-D-10-05030.1>
- Pearlman J, Bushnell M, Coppola L, Karstensen J, Buttigieg PL, Pearlman F, Simpson P, Barbier M, Muller-Karger FE, Munoz-Mas C, Pissierssens P, Chandler C, Hermes J, Heslop E, Jenkyns R, Achterberg EP, Bensi M, Bittig HC, Blandin J, Bosch J, Bourles B, Bozzano R, Buck JJH, Burger EF, Cano D, Cardin V, Llorens MC, Cianca A, Chen H, Cusack C, Delory E, Garello R, Giovanetti G, Harscoat V, Hartman S, Heitsenrether R, Jirka S, Lara-Lopez A, Lantéri N, Leadbetter A, Manzella G, Maso J, McCurdy A, Moussat E, Ntoumas M, Pensieri S, Petihakis G, Pinardi N, Pouliquen S, Przeslawski R, Roden NP, Silke J, Tamburri MN, Tang H, Tanhua T, Telszewski M, Testor P, Thomas J, Waldmann C and Whoriskey F (2019) Evolving and Sustaining Ocean Best Practices and Standards for the Next Decade. *Front. Mar. Sci.* 6:277. doi: 10.3389/fmars.2019.00277
- Snowden D, Tsontos VM, Handegard NO, Zarate M, O' Brien K, Casey KS, Smith N, Sagen H, Bailey K, Lewis MN and Arms SC (2019) Data Interoperability Between Elements of the Global Ocean Observing System. *Front. Mar. Sci.* 6:442. doi: 10.3389/fmars.2019.00442
- Stommel, H., 1989: The Slocum Mission. *Oceanography*, 2(1), pp. 22–25.
- Tanhua T, Pouliquen S, Hausman J, O'Brien K, Bricher P, de Bruin T, Buck JJH, Burger EF, Carval T, Casey KS, Diggs S, Giorgetti A, Graves H, Harscoat V, Kinkade D, Muelbert JH, Novellino A, Pfeil B, Pulsifer PL, Van de Putte A, Robinson E, Schaap D, Smirnov A, Smith N, Snowden D, Spears T, Stall S, Tacoma M, Thijsse P, Tronstad S, Vandenberghe T, Wengren M, Wyborn L and Zhao Z (2019) Ocean FAIR Data Services. *Front. Mar. Sci.* 6:440. doi: 10.3389/fmars.2019.00440
- Testor, P. & Co_Authors (2010). "Gliders as a Component of Future Observing Systems" in *Proceedings of OceanObs'09: Sustained Ocean Observations and Information for Society (Vol. 2)*, Venice, Italy, 21-25 September 2009, Hall, J., Harrison, D.E. & Stammer, D., Eds., ESA Publication WPP-306, doi:10.5270/OceanObs09.cwp.89
- Testor, P., et al. (2019), OceanGliders: a component of the integrated GOOS, *Front. Mar. Sci.*, 6:422, <http://doi.org/10.3389/fmars.2019.00422>.

6. List of Acronyms

| Acronym | Definition |
|------------|--|
| ACDD | Attribute Convention for Data Discovery |
| ANFOG | Australian National Facility for Ocean Gliders |
| AODN | Australian Ocean Data Network |
| API | Application Programming Interface |
| AUV | Autonomous Underwater Vehicle |
| BODC | British Oceanographic Data Centre |
| BUFR | Binary Universal Form for the Representation of meteorological data |
| CDI | Common Data Index (SeaDataNet catalogue) |
| CF | Climate and Forecast (Convention) |
| CMEMS | Copernicus Marine Environmental Monitoring Service |
| COARDS | Cooperative Ocean/Atmosphere Research Data Service |
| CSR | Cruise Summary Report (SeaDataNet catalogue) |
| DAC | Data Assembly Centre |
| DG MARE | Directorate-General for Maritime Affairs and Fisheries |
| DM | Delayed Mode |
| DMQC | Delayed Mode Quality Control |
| EC | European Commission |
| EDIOS | European Directory of the Ocean Observing System (SeaDataNet catalogue) |
| EDMED | European Directory of Marine Environmental Data (SeaDataNet catalogue) |
| EDMERP | European Directory of Marine Environmental Research Projects (SeaDataNet catalogue) |
| EDMO | European Directory of Marine Organisations (SeaDataNet catalogue) |
| EEA | European Environment Agency |
| EGO | Everyone's Gliding Observatories |
| EMODnet | European Marine Observation and Data Network |
| EOOS | European Ocean Observing System |
| EOSC | European Open Science Cloud |
| ECV | Essential Climate Variables |
| EOV | Essential Ocean Variables |
| EU | European Union |
| EuroGOOS | European Global Ocean Observing System |
| FAIR | Findable-Accessible-Interoperable-Reusable |
| FTP | File Transfer Protocol |
| GDAC | Global Data Assembly Centre |
| GOOS | Global Ocean Observing System |
| GRA | GOOS Regional Alliances |
| GROOM | Gliders for Research, Ocean Observation and Management |
| GTS | Global Telecommunication System |
| GTSP | Global Temperature and Salinity Profile Programme |
| HELCOM | Baltic Marine Environment Protection Commission – Helsinki Commission |
| HPC | High Performance Computing |
| ICES | International Council for the Exploration of the Sea |
| IMOS | Integrated Marine Observing System |
| IOC-UNESCO | Intergovernmental Oceanographic Commission – United Nations Educational, Scientific and Cultural Organization |
| IODE | International Oceanographic Data and Information Exchange |
| IOOS | Integrated Ocean Observing System |
| ISO | International Organization for Standardization |
| JCOMMOPS | Joint Technical Commission for Oceanography and Marine Meteorology in situ Observations Programme Support Centre |



sdn-userdesk@seadatanet.org – www.seadatanet.org

SeaDataCloud - Further developing the pan-European infrastructure for marine and ocean data management

Grant Agreement Number: 730960

| | |
|----------|--|
| MoU | Memorandum of Understanding |
| MSFD | Marine Strategy Framework Directive |
| NetCDF | Network Common Data Form |
| NOAA | National Oceanic and Atmospheric Administration |
| NODC | National Oceanographic Data Centre |
| NRT | Near Real Time |
| NVS | NERC (Natural Environment Research Council) Vocabulary Server |
| ODIP | Ocean Data Interoperability Platform |
| ODV | Ocean Data View |
| OG1.0 | OceanGliders data exchange format 1.0 |
| OGC | Open Geospatial Consortium |
| OGC-SWE | Open Geospatial Consortium – Sensor Web Enablement |
| OGDMTT | OceanGliders Data Management Task Team |
| OSPAR | Oslo Paris Convention (Convention for the Protection on the Marine Environment of the North-East Atlantic) |
| PI | Principal Investigator |
| QA | Quality Assurance |
| QC | Quality Control |
| REC | Recovery (data version obtained from Slocum glider platform upon recovery) |
| ROOS | Regional Operational Oceanographic Systems |
| RTQC | Real Time Quality Control |
| SensorML | Sensor Model Language |
| SOCIB | Sistema d'observació i predicció costaner de les Illes Balears (Balearic Islands Coastal Observing and Forecasting System) |
| SOS | Sensor Observation Service |
| SSN | Semantic Sensor Network (Ontology) |
| SWE | Sensor Web Enablement |
| TAC | Thematic Assembly Centre |
| TESAC | Temperature, Salinity and Current (report) |
| THREDDS | Thematic Real-time Environmental Distributed Data Services |
| UEA | University of East Anglia |
| UG2 | Underwater Glider User Group |
| UN | United Nations |
| UNESCO | United Nations Educational, Scientific and Cultural Organization |
| W3C | World Wide Web Consortium |
| WMO | World Meteorological Organization |