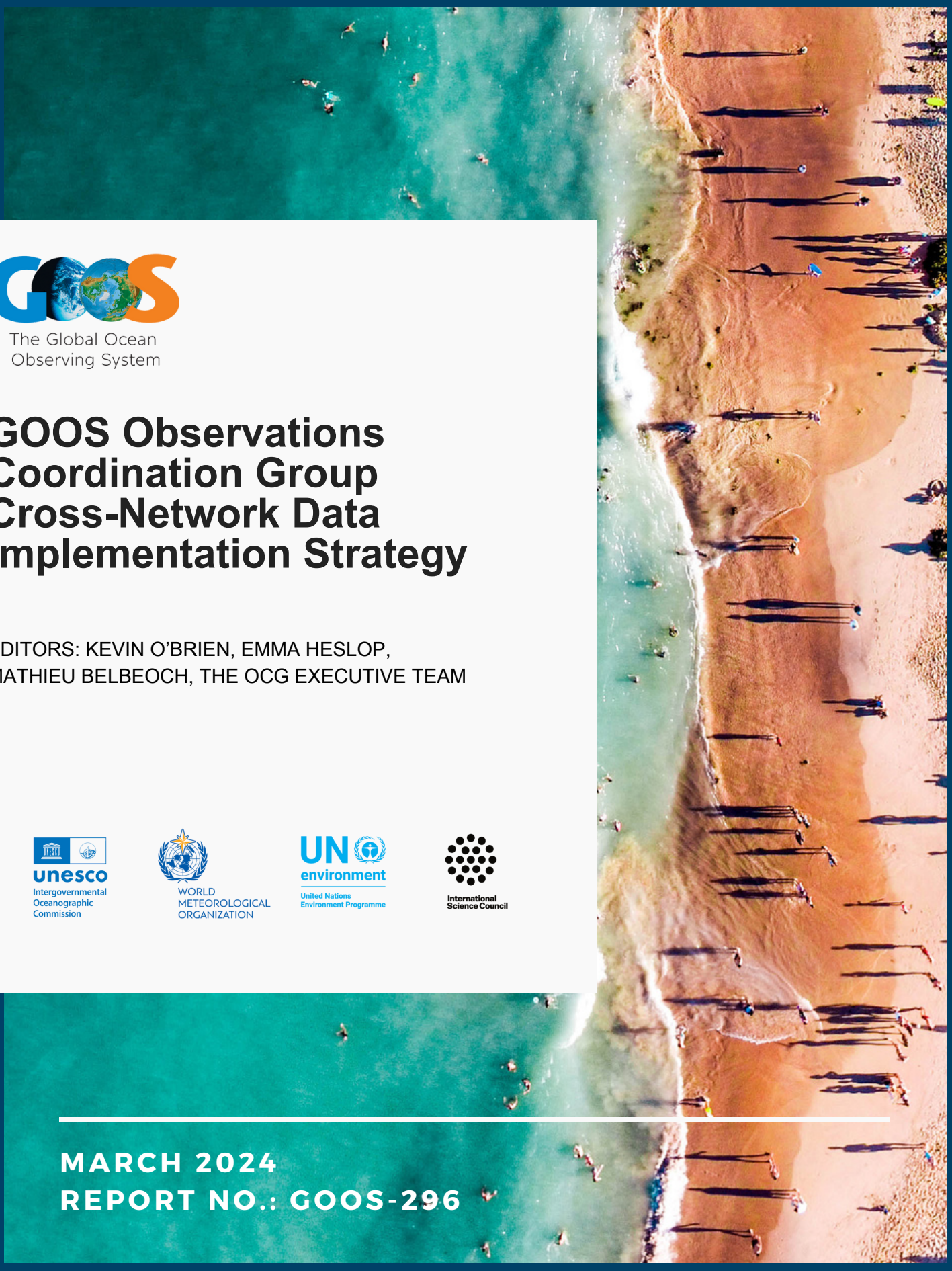
GOOS PUBLICATION



The Global Ocean Observing System

GOOS Observations Coordination Group Cross-Network Data Implementation Strategy

EDITORS: KEVIN O'BRIEN, EMMA HESLOP, MATHIEU BELBEOCH, THE OCG EXECUTIVE TEAM

unesco Intergovernmental Oceanographic Commission

WORLD METEOROLOGICAL ORGANIZATION

UN environment United Nations Environment Programme

International Science Council

MARCH 2024
REPORT NO.: GOOS-296

Table of Contents

I. Introduction

Background

Data management and data flows are key components of both the Global Ocean Observing System (GOOS) 2030 Strategy and its Roadmap. The GOOS 2030 Strategy[1] is a "call to action for ocean observation" and, in combination with the GOOS 2030 Strategy Roadmap[2], looks to define a path forward to meet increased societal needs for a fully integrated global ocean observing system that will provide the critical ocean information required for policymakers and the public and private sectors to adapt to and mitigate climate change, protect ocean health and support sustainable growth as well as marine-based activities. This path includes improving the data flow and interoperability among the global in situ ocean observing networks.

The GOOS Observations Coordination Group (OCG) works to coordinate the activities of the active global ocean observing networks to ensure an effective and integrated global ocean observing system. One of the eight strategic areas of work for the OCG is data management and the OCG is committed to improving data interoperability, both within the global networks as well as to support service delivery with our close partners, such as the World Meteorological Organization (WMO) and the International Oceanographic Data and Information Exchange (IODE) of IOC, as well as external stakeholders.

The ocean observing networks operating under the umbrella of the OCG are the backbone of the global ocean observing system, and deliver hundreds of thousands of observations every day. These networks are heterogeneous and operate different platforms on varying space and timescales collecting multiple Essential Ocean Variables[3] (EOVs). Yet, despite this heterogeneity, one of the core Strategic Objectives of the GOOS 2030 Strategy is for all ocean data under GOOS to be FAIR (findable, accessible, interoperable and reusable) and to reach its users with appropriate latency and quality. In order to reduce friction to data flow, enhance and track FAIR data implementation, and support the ocean observing component of the ocean data infrastructure, the OCG Data Management Team[4], in coordination with the global ocean observing networks and key stakeholders, has developed the GOOS OCG Cross-Network Data Implementation Strategy.

---

[1] https://www.goosocean.org/index.php?option=com_oe&task=viewDocumentRecord&docID=24590

[2] https://www.goosocean.org/index.php?option=com_oe&task=viewDocumentRecord&docID=26687

[3] https://goosocean.org/what-we-do/framework/essential-ocean-variables/

[4] https://goosocean.org/who-we-are/observations-coordination-group/data-management/

Developing the Data Implementation Strategy

In order to better understand the current state of the data management infrastructure that underpins the global ocean observing system, the OCG Data Management Team conducted a data interoperability analysis of the OCG global ocean observing networks to map the data flows, for real time data, delayed mode data , and metadata, across all 13 global networks and 84 national contributors, and through this identified opportunities for improvement. This process illustrated the complex and, at times, chaotic nature of these systems (see Appendix B for more information in the mapping). The results of the data flow mapping, along with careful examination of existing data strategies from the WMO, UN Ocean Decade and IODE, was critical to the development of this OCG Cross-Network Data Implementation Strategy. The Strategy is based around the FAIR (Findable, Accessible, Interoperable and Reusable) data principles supporting free and open data and provides specific implementation requirements for the global networks to meet these data goals.

As the complexity of the data landscape continues to grow, for example with the emergence of AI, cloud computing and digital twins, it is more important than ever to enable the trusted exchange of data and information through interoperable, accessible and timely data services. Although the ocean community has galvanized around the FAIR data principles for data exchange, more specific requirements are needed in order to actually achieve interoperability. In addition, issues with data discovery continue to impede users' ability to find and use the data they want or need. Rather than yet another data strategy, the GOOS OCG decided to provide an implementation strategy which would align with existing strategies and point the way to achieving the stated goals of all of those strategies: improved interoperability, discoverability and usability of ocean data and information. This Strategy will be reviewed and updated as needed to enable agility in the face of technological and data management innovations.

II. OCG Data Implementation Strategy Vision, Goals and Objectives

Vision

To provide frictionless, integrated, interoperable, and FAIR-compliant access to ocean data and metadata from across the OCG global ocean observing networks, and to ensure that all data reaches a uniform end point to fully support future federated data systems. To continually improve seamless access and delivery efficiency through innovation, stakeholder dialogue, and engagement with international ocean data communities.

Goals

- Integration of data and metadata across OCG networks
- FAIR compliance to improve discovery and use of metadata, data and data services across all OCG networks

Objectives

- Provide access to data endpoints (i.e., global repositories) for each OCG network which can be federated through common middleware (such as ERDDAP)
- Ensure high quality data are freely and openly available in near real time from the WMO Global Telecommunication System (GTS) and/or other data access services
- Ensure high quality data/metadata are discoverable and harvestable
- Ensure high quality data/metadata are available through identified global repositories
- Ensure data[5] are fully documented and required metadata is available through OceanOPS
- Ensure data/metadata are properly archived and citable to maximize reuse

III. Key Implementation Elements

ERDDAP

Through experimentation and various pilot efforts, OCG has identified the data platform ERDDAP[6] as the key to improving interoperability and meeting FAIR compliance guidelines within OCG and the global ocean community. ERDDAP is an open source data broker that is widely used within the marine meteorological and oceanographic community to document, serve and archive data. One of the strengths of ERDDAP is its

[5] OceanOPS is the WMO/IOC Operational Centre, which, as part of the OCG, sets minimum metadata requirements, tracks the delivery of observations from the global networks through the metadata, and also enables visualizations and analysis of the observing system status

[6] https://coastwatch.pfeg.noaa.gov/erddap/

ability to interact with many different formats of data. This capability allows communities to work in the format with which they are most familiar while sharing the exact same data to other communities in the formats they are most familiar with. In a sense, it can be considered an on-the-fly data translator. This ability to abstract users and producers from data formats is a huge advantage.

In addition, ERDDAP is capable of machine-to-machine interactions, as well as manual interactions. This means that users can easily develop programmatic access for ERDDAP data and metadata. Automatic metadata translation is also done on the fly, so that it is possible to harvest metadata in the ISO-19115[7] standard, making it very straightforward for harvesting engines. Included in this could be data and/or metadata harvests to support the WMO Integrated Global Observing System (WIGOS), discovery tools, and the Intergovernmental Oceanographic Commission (IOC) Ocean Data and Information System (ODIS), as well as to integrate seamlessly with the future WMO Information System (WIS) 2.0 decentralized GTS evolution. ERDDAP also supports, natively, schema.org[8] markup language, which creates structured data in support of the semantic web to make metadata more readable by search engines such as Google, etc. There is no extra effort required for this to happen.

Finally, in the OCG Data Implementation Strategy, ERDDAP is identified as the key uniform endpoint data service that will ensure OCG network data is accessible and suitable for federation across the distributed data end points.

Requirements

To consistently meet the objectives across the different networks and to track achievement and delivery, the global ocean observing networks are encouraged to work towards meeting all of the following data and metadata requirements, and to continue maturing in these areas.

The OCG will actively work to support and assist networks in achieving these data requirements as far as is practical. Networks that have achieved these requirements will be considered mature for data and metadata (see OCG Network Attributes[9]).

---

[7] https://www.iso.org/standard/53798.html

[8] https://schema.org/

[9] https://www.goosocean.org/index.php?option=com_oe&task=viewDocumentRecord&docID=33315

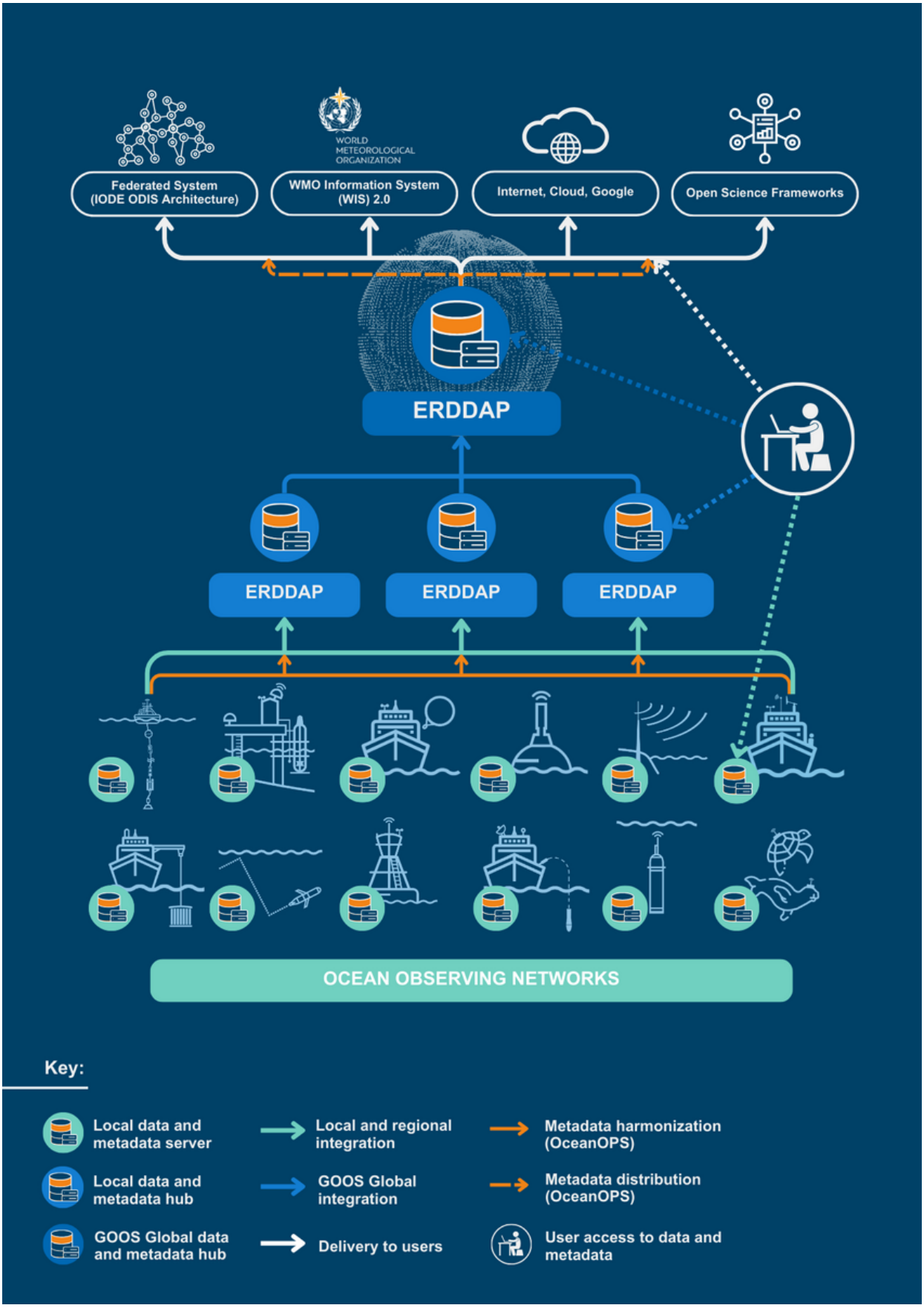


Figure 1. *A schematic of the proposed GOOS OCG Federated data system which provides local, regional and global access to distributed data and metadata resources.*

Table 1. *GOOS Observations Coordination Group Data Implementation Requirements*

Real Time Data	
OCG-R1	Data shall be exchanged in real time (with minimum delay) via the WIS/GTS of the WMO in approved formats/templates.
OCG-R2	Data shall be available in real time or near-real time on the Internet through interoperable services (preferably ERDDAP) freely and without any restriction. Community agreed quality control procedures shall be applied in real-time and adjusted values made available when possible.
Delayed Mode Data	
OCG-R3	Each network shall have at least one identified global data repository. This Global Data Repository may be one or multiple (mirrored) repositories, or they may be data endpoints that can be federated into a virtual global repository.
OCG-R4	Data and data products shall be available through publicly accessible ERDDAP services. These distributed ERDDAP services will be federated under a single OCG ERDDAP focal point.
OCG-R5	NetCDF is the preferred data file format, though ERDDAP services can act as a data format translator if needed.
OCG-R6	Additional platform metadata should be available through the Global Data Repository and harvestable by machine-2-machine services.
Metadata	
OCG-R7	Networks shall have a defined uniform metadata content that includes at least the minimum OceanOPS requirements, thereby ensuring that they are compliant with the WIGOS metadata requirements.[10] Note that OceanOPS is the authoritative source through which WIGOS metadata are submitted to OSCAR[11] for all oceanographic and marine meteorological platforms.
OCG-R8	Discovery and Use metadata shall be based upon a well-documented community standard, including a persistent and unique WMO/WIGOS identifier allocated by OceanOPS and use controlled vocabularies. Examples of such community standards include the Climate and Forecast Metadata conventions[12], the WIGOS metadata standard[3], the Darwin Core standard[13] for biological data and Seadatanet vocabularies[14].
OCG-R9	Platform and Discovery metadata shall be exchanged with OceanOPS utilizing machine-2-machine services and avoiding multiple redundant manual transmissions.
Best Practices	
OCG-R10	Each network should have an active data team that will help implement the OCG data requirements.
OCG-R11	Each network should have identified best practices on data infrastructure, workflows and quality control processes.
OCG-R12	Raw/real-time data, delayed mode data and data products should be archived and have unique identifiers created (i.e., Digital Object Identifier (DOI)) for citation and reuse.

[10] https://library.wmo.int/doc_num.php?explnum_id=10109
[11] https://www.goosocean.org/index.php?option=com_oe&task=viewDocumentRecord&docID=21727
[12] http://cfconventions.org/
[13] https://dwc.tdwg.org/
[14] https://seadatanet.org/Standards/Common-Vocabularies

Tracking Progress

The recommendations provided by this Strategy are meant to be specific and actionable. In order to track compliance with the recommendations, they must also be measurable. OceanOPS currently produces an annual assessment of network capabilities in the GOOS Ocean Observing System Report Card[15], OCG will continue to track achievement of the network data requirements through the Report Card, and so track improvement in the overall OCG data ecosystem. Figure 2 illustrates the relevant data categories where OCG tracks each network's progress towards improved data practices. Moving forward these scores will be based upon the requirements outlined above, and as a part of a broader matrix that OCG is developing to track maturity levels across the global ocean observing networks. Data, metadata and best practices are key components in determining maturity levels.

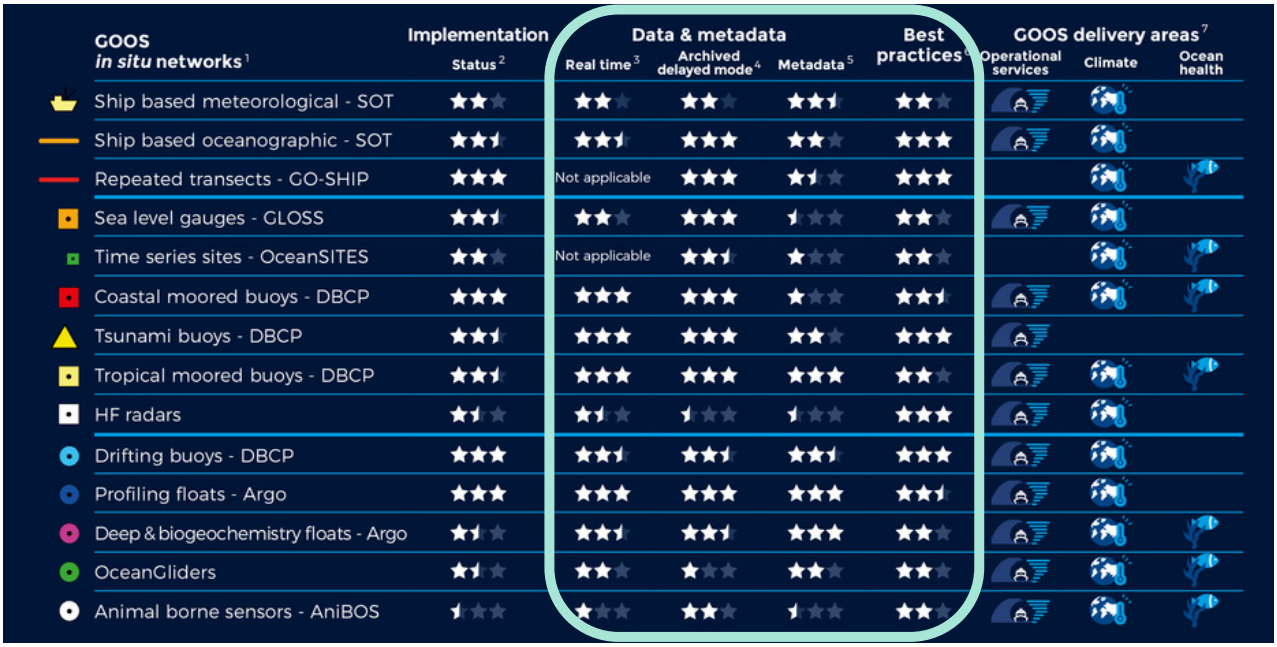


Figure 2. *From the GOOS Ocean Observing System Report Card 2023. The data, metadata and best practice scores are highlighted.*

[15] https://www.ocean-ops.org/reportcard/

Links to the Ocean Data Community

The development of this data implementation strategy was closely coordinated with complementary efforts in the global community, including:

1. WMO Unified Data Policy[16]
2. IOC Strategic Plan for Ocean Data and Information Management (2023-2029)[17]
3. Development of the IODE Ocean Data Information System and Ocean InfoHub[18]
4. UN Ocean Decade Data Strategy and Information Strategy[19]
5. OceanOPS 5 Year Strategic Plan[20]

One goal of the OCG Data Implementation Strategy is to ensure that when data complies with the Strategy requirements, it will also comply with the requirements from these external data management strategies. This is important to ensure that OCG network-collected data is available to support the WMO and IOC stakeholders without adding extra burden upon the OCG network data producers.The OCG Data Management Team will continue to engage with these key international stakeholders and monitor the evolution of their data strategies. It is important to note that the OCG Data Team and OceanOPS stand ready to assist networks in the implementation of these data recommendations.

OceanOPS is a key component in this connection to external stakeholders, in particular the WMO and IODE. OceanOPS is the metadata repository responsible for routinely populating discovery and access systems though the WMO (OSCAR, WIGOS), the IOC (IODE ODIS and OIH) and the UN Ocean Decade (DiTTO, etc). OSCAR metadata are the foundation of a number of gap analysis tools and processes (GBON, RRR). This connection is why it is so critical that OCG network metadata is complete and up-to-date in the OceanOPS system.

[16] https://library.wmo.int/records/item/58009-wmo-unified-data-policy

[17] https://unesdoc.unesco.org/ark:/48223/pf0000385113

[18] https://odis.iode.org/

[19] https://unesdoc.unesco.org/ark:/48223/pf0000385542

[20] https://www.goosocean.org/index.php?option=com_oe&task=viewDocumentRecord&docID=21727

IV. Next Steps

The OCG Cross-Network Data Implementation Strategy is an ambitious effort to transform the data management framework of the GOOS global networks and also to provide a blueprint for how other projects/programs can improve the interoperability and usability of their own data.
Wherever there are gaps or improvements needed in the data value chain, the OCG will endeavor to support its networks in seeking the resources to advance towards the ultimate goal of frictionless and FAIR data delivered from across all networks. The OCG Data Implementation Strategy also adds clarity to the sometimes vague notion of FAIR data by providing concrete and actionable methods to achieve such compliance.

The implementation of the OCG data requirements will depend heavily on the data teams of each network. The OCG Exec, through the activities of the Data Vice-Chair, and OceanOPS, will support or help find support for enhancements to the data structures of the OCG networks, OceanOPS, etc. This includes support for ERDDAP installation, configuration and usage, as well as support for the implementation of metadata services to support OceanOPS. This support will be through quarterly OCG Data and Metadata Roundtables, direct interaction with network data teams, and attendance at relevant network workshops/meetings.

The data flow mappings have provided a benchmark for the current status of the OCG data infrastructure and the OCG annual meeting, along with the annual GOOS Ocean Observing System Report Card, will provide a way of reporting progress towards full implementation of the OCG Data Implementation Strategy.

Achieving these data goals will have a significant impact on the availability and useability of the data generated by the OCG global networks and serve to maximize the value of that data by reaching as many users as possible in easy to use and understandable services and formats. The OCG will annually assess progress towards the overall data management goals and adjust as necessary to reach them.

APPENDIX A. The FAIR data Principles

In 2016, Scientific Data published the FAIR Guiding Principles for scientific data management and stewardship[21]. The goal was to establish a set of guidelines to improve data sharing between and amongst communities.

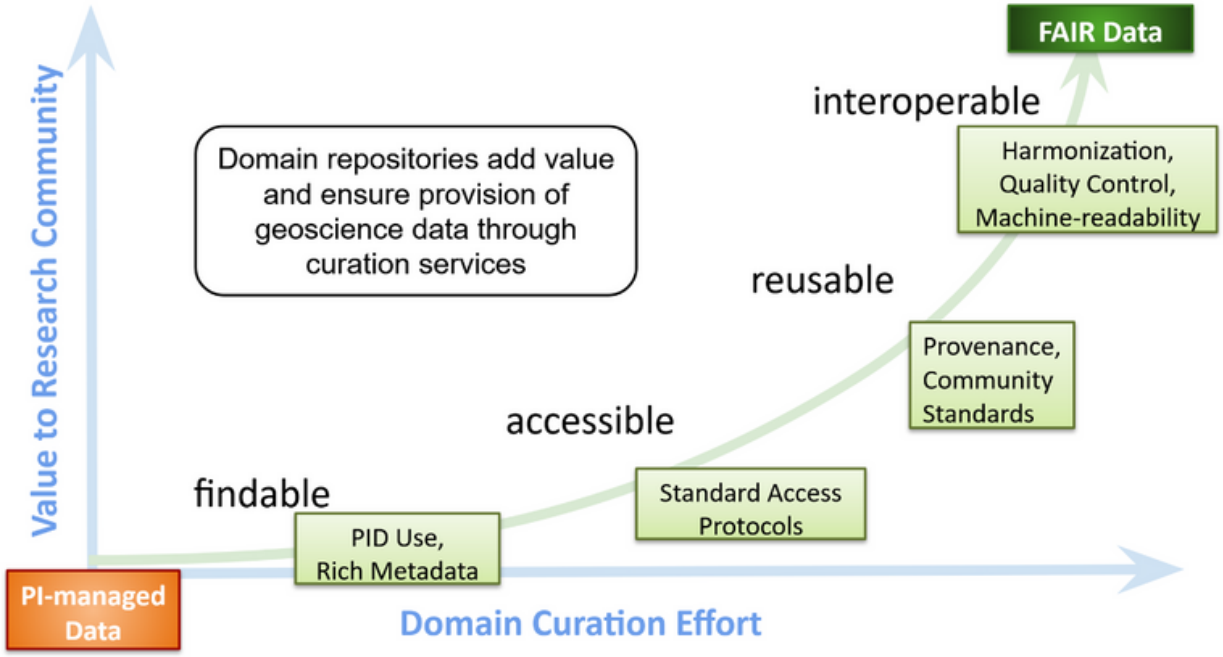


Figure 3. *Increasing data value through compliance with FAIR principles – from Soenen, K. (2021). Reusing Open Data with ERDDAP and Python. https://doi.org/10.5281/zenodo.5684719*

There are four guiding Principles:

1. Findable
   - The first step in using data is to find them. Metadata and data should be easy to find for both humans and computers. Machine-readable metadata are essential for automatic discovery of datasets and services, so this is an essential component
2. Accessible
   - Once the user finds the data or metadata, they must be able to have free and open access using standard protocols that support machine-to-machine interactions

[21] Wilkinson, M., Dumontier, M., Aalbersberg, I. et al. The FAIR Guiding Principles for scientific data management and stewardship. Sci Data 3, 160018 (2016). https://doi.org/10.1038/sdata.2016.18

3. Interoperable
    - To be interoperable, metadata must use standard vocabularies that follow the FAIR principles
4. Reusable
    - The ultimate goal of FAIR is to optimize the reuse of data. To achieve this, metadata and data should be well-described so that they can be replicated and/or combined in different settings.

The four principles do provide a guide for improving interoperability, but it is a rather vague guide, and it can be difficult to define actionable ways forward. The OCG data implementation strategy is an attempt to provide clear guidance on how to comply with the FAIR principles.

APPENDIX B. OCG Data Flow Mapping

Introduction

As part of the Observations Coordination Group (OCG) Data Strategy, and Metadata harmonization efforts, and in line with a request from the tenth GOOS Steering Committee meeting in September 2020, GOOS initiated a Data and Metadata mapping project to clearly map out the data and metadata flows for both real-time and delayed mode data, across all the recognized OCG global networks: Argo, OceanGliders, Data Buoy Cooperation Panel (DBCP), Ship Observations Team (Voluntary Observing Ships (VOS), Ships of Opportunity Program (SOOP), Automated Shipboard Aerological Programme (ASAP)), the Global Ocean Ship-Based Hydrographic Investigations Program (GO-SHIP), OceanSITES, Global Sea Level Observing System (GLOSS), Animal Borne Ocean Sensors (AniBOS), and High Frequency (HF) Radar.

Our aims for this mapping exercise were to: 1) enable those outside the networks to better understand how network data moves through the global and national data management systems; 2) enable us in identifying gaps and areas where we can potentially improve or better support data and metadata access; 3) ensure that the required metadata is accessible and flowing into the OceanOPS monitoring system; 4) use as a base for a cross-network data strategy, ensuring that the data from global networks reaches existing and future global access points for both operational real-time and quality controlled delayed mode data. This is a focused effort to increase value and visibility of OCG network data by improving its interoperability and ensuring that OCG data meets its critical role as an integral part of the global ocean information digital ecosystem.

The complete set of data mappings is available here.[22] The information and insight from this data mapping informed and guided the development of an OCG Data Implementation Strategy, to better support observational network development in these areas, and recommend best practices to the community. Understanding these data and metadata flows and identifying areas of enhancement are crucial in order to increase FAIR compliance of OCG network data and compliance with the WMO Unified Data Policy and IOC Data Policy and Terms of Use (2023).

The data mapping contains data structures, QC elements, and key performance indicators such as data availability, timeliness, and completeness of metadata, to identify the current state of data and metadata flows for the networks. With this picture of the data flows within the different networks, we can also work with WMO and IODE to integrate this information into their data mapping efforts to extend mapping beyond the scope of just OCG and its networks.

Beyond the work of the GOOS OCG the data maps provided here are useful for global, national and regional observing and data management systems, GOOS partners and others that want to understand the existing data pathways.

[22] https://goosocean.org/index.php?option=com_oe&task=viewDocumentRecord&docID=31176

ACKNOWLEDGMENTS

The Global Ocean Observing System

unesco Intergovernmental Oceanographic Commission

WORLD METEOROLOGICAL ORGANIZATION

UN environment United Nations Environment Programme

International Science Council

Published in 2024 by the Global Ocean Observing System under the Intergovernmental Oceanographic Commission of UNESCO, 7, Place de Fontenoy, 75352 Paris 07 SP, France

Editorial Team: Kevin O'Brien, Emma Heslop, Mathieu Belbeoch, The OCG Executive Team

For bibliographic purposes, this publication should be cited as follows : Observations Coordination Group (OCG) Data Implementation Strategy (2024), GOOS Report No.296

Contact: Kevin.M.O'Brien@noaa.gov