**9th Session of the JCOMM Observations Coordination Group**

**14 - 17th May 2018, Brest, France**

**Report Title: OCG Data Management Strategy and Update on Data Integration Activities**

**Authors: Kevin O’Brien**

**Agenda Item Reference: 9.1**

**Date: May 18, 2017**

**Draft: version 1**

**1. SUMMARY**

This report will have two parts. The first part is to initiate the discussions for formalizing the OCG Data Management Strategy and the second part is a discussion of currently activities within OCG to improve integration and interoperability of data from OCG networks.

**2. REPORT CONTENT**

OCG Data Management Strategy

Recently approved at JCOMM V, the OCG has elected an OCG Data and Information Vice-Chairperson. The role of this position will be to coordinate data management activities of interest to the Observations Programme Area (OPA) of JCOMM, and particularly the OPA response to the Joint WMO and IOC Strategy for Marine Meteorological and Oceanographic Data Management (2018-2021); including promoting adoption of consistent standards and practices for oceanographic and marine meteorological observations data sharing and exchange; and identify and encourage development of integrated and interoperable solutions in this regard for the benefit of JCOMM stakeholders. This shall be done in coordination with the JCOMM Data Management Coordination Group (DMCG)[[1]](#footnote-1).

Links to JCOMM DM Strategy

OCG has provided input to the Joint WMO-IOC Strategy for Data Management (2018-2021) document. OCG will either lead or be involved in several activities described by this document. These outcomes and activities include:

* Promoting data sharing in compliance with WMO Resolution 40, WMO Resolution 60 and tie IOC Oceanographic Data Exchange policy
* Achieving a more comprehensive and standardized collection of ocean and marine met data in real time and near real time
* Integration of ocean and marine met data, including dataflow, archival and enhanced data access.

Open Access to GTS pilot project

OCG is happy to report on the successful pilot project aimed at providing an easier workflow for data providers to put data onto the GTS, and for data users to get data from the GTS. OCG worked with five different organizations to put ocean observation onto the GTS, using the improved workflow. Here is a list of platforms that provided data to the GTS:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **WMO ID** | **Platform Name** | **Platform Type** | **Institution** | **Location (approx)** |
| 6101404 | Algeciras | Coastal Mooring | PdE | Med Sea |
| 6201070 | Langosteira | Coastal Mooring | PdE | Atlantic (Spanish coast) |
| 6101403 | E2M3A | E2M3A | OGS | Adriatic Sea |
| 4802011 | 1001 | Saildrone | PMEL | Arctic |
| 4802012 | 1002 | Saildrone | PMEL | Arctic |
| 4802013 | 1003 | Saildrone | PMEL | Arctic |
| 5801950 | 1005 | Saildrone | PMEL | TPOS |
| 5801951 | 1006 | Saildrone | PMEL | TPOS |
| 5100275 | Aloha Cabled Observatory | Bottom Pressure Recorder | PacIOOS | Hawaii |
| Several |  | Ships | Volvo Ocean Race | Global |

The result of the pilot project was presented at the JCOMM V Marine Technology meeting in Geneva, October 2017. It was recommended at that meeting that the Open Access to GTS pilot should continue, though there may be obstacles to that happening. This will be discussed further in the “Recommendations” section. However, though there was some hesitation at the JCOMM session about embracing such a new direction, there was also much interest from those who are providing, or would like to provide, their data to the GTS. At a minimum, there is clearly a need for an improved process for accessing the GTS.

Figure 1 Data sent to the GTS from Saildrones in Tropical Pacific

An additional aspect of the Open Access to GTS work was a GTS and CMEMS data comparison done by Simona Simoncelli and Alessandro Grandi, from INGV. Their findings indicated the importance of accessing GTS data, as most data from GTS were available with 12 hours, whereas data from CMEMS was only available after a 24 hour period. They also discovered issues with data availability, both in the GTS and CMEMS, which needs to be studied further.

Data Integration progress

OCG is happy to report the following progress in terms of data integration:

* Global Drifter Program

AOML GDP Director, Rick Lumpkin, and OCG member O’Brien submitted a proposal to the NOAA Big Earth Data Initiative to secure funding to implement an improved workflow for GDP quality controlled, interpolated data. This proposal, which runs for one year, was accepted and a small amount of funds approved. Through this work, AOML is able to modernize their workflow to provide GDP delayed mode through standard formats, ingest that data into an ERDDAP server to provide interoperable access, and create archive packages acceptable to the US National Archive center, NCEI. This work builds upon previous OCG work to integrated GDP delayed mode data into ERDDAP services, and provides an improved workflow for the AOML GDP program

* Marine Explorers from Pole To Pole (MEOP)

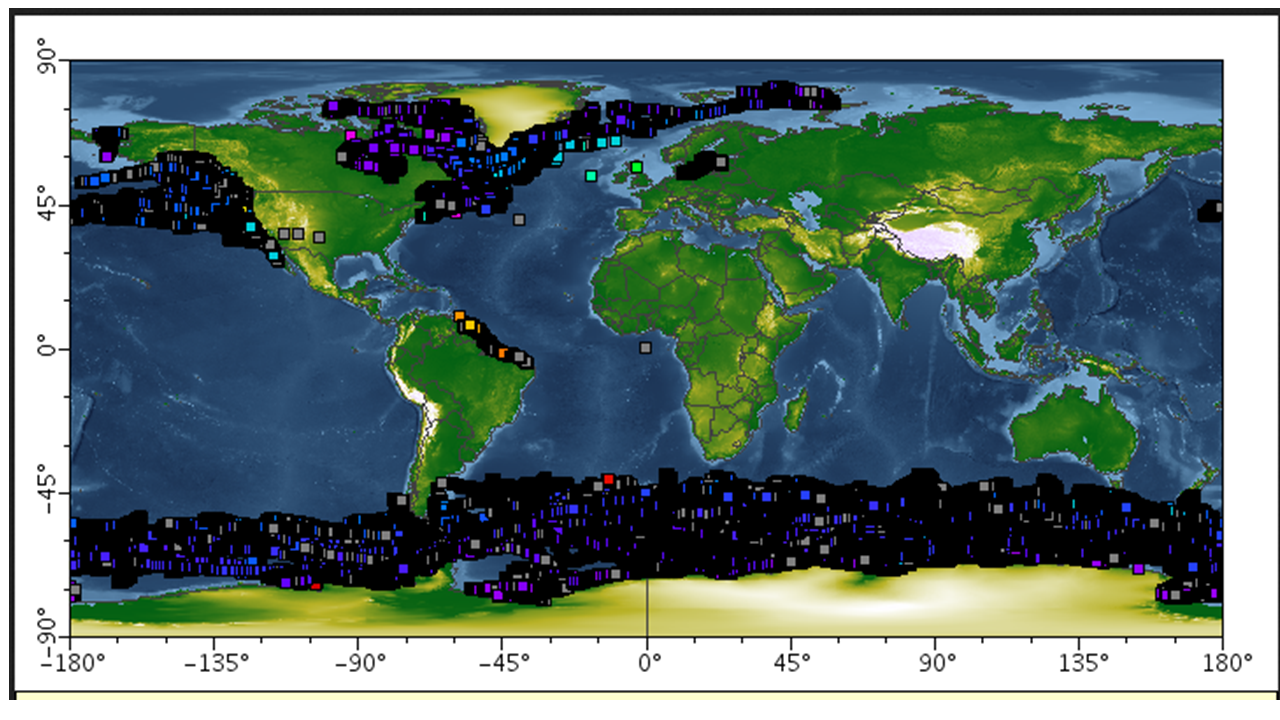
OCG has integrated the MEOP data, previously available through an email request and Google drive files, into an ERDDAP service and made these publically available. The integration of these data into an ERDDAP services has allowed us to provide a view of the entire collection of data through a single URL. We will discuss next steps with the data from this emerging network at the OCG meeting. This service was used to help identify the “platform” reports for the JCOMMOPS report card.

Figure 2 Overview of all profiles from MEOP database available through ERDDAP services

* SAMOS

OCG has begun working with the SAMOS shipboard data to load this data into an ERDDAP service. Though the data is in a THREDDS Data Server (TDS) currently, this service has the disadvantage of not allowing a user to view the whole collection of data, but rather subsets both by vessel and time.

* EGO Gliders/SOCIB

As described above with SAMOS, OCG has begun working with the glider data from the EGO project, using the NetCDF files from their ftp archive. Personal communication from Pierre Tester to O’Brien at EGU in April, 2018, confirmed which the appropriate files to use for this activity were. Testor also mentioned that the SOCIB gliders, from Puerto del Estado, were in the process of finalizing their NetCDF format to match the EGO format. Once that process is completed, OCG will include the SOCIB files as well.

* Connection with IMOS

IMOS AODN Director Roger Proctor and O’Brien met face-to-face at the EGU meeting in April 2018. There they discussed the use of ERDDAP as an additional access method to the current IMOS data services. O’Brien agreed to assist IMOS in configuring an ERDDAP service using example files from IMOS. This work will continue in the coming year.

* Cloud Services

There has been some interest from cloud services, particularly Amazon Web Services and Google Cloud, in the real-time ocean data that is collected and integrated into ERDDAP services by the Observing System Monitoring Center.

Finally, it is worth noting that the last year brought about large-scale integration of ERDDAP as a data service of choice across the international community. Those embracing ERDDAP include EMODnet Physics, which now has an ERDDP server with over 50,000 datasets available. The Marine Institute in Galway continues to rely on ERDDAP as a data service of choice to provide data and information through their Irish Digital Ocean. In addition, US national services, such as U.S. IOOS and NOAA’s National Center for Environmental Information (NCEI) are already, or will be soon, using ERDDAP as an integral part of the data access services they provided. ERDDAP support is as strong as ever, as it reaches its 10th birthday in 2018. The upcoming release of ERDDAP will also include the ability to ingest data through RESTful/machine-to-machine services thereby providing a key capability allowing ERDDAP to integrate into automated data management processes.

**3. DECISIONS, ACTIONS and RECOMMENDATIONS**

Open Access to GTS

As mentioned above, the pilot project was well received, especially by those interested and having difficulties putting data onto the GTS. One issue that needs to be resolved is that of finding willing partners who 1) have access for sending data to the GTS and 2) are willing to embrace the workflow established by the pilot project. US National Data Buoy Center (NDGC) was instrumental in putting data onto the GTS during the duration of the pilot project. However, they would prefer to end the pilot project immediately, and concentrate on what resources are required to move this into operations. There is a meeting planned at NDBC for mid-July to address these issues. However, NDBC is not prepared to be the focal point for an international effort.

O’Brien was able to discuss this project at the EGU meeting in April 2018 and identified several potential partners that have access to the GTS and may be willing to partner in this effort. These institutions are:

* Integrated Marine Observing System (IMOS), Australia
* Marine Institute, Ireland
* British Oceanographic Data Center (BODC), UK
* Copernicus Marine Services

One recommended strategy for moving forward would be to focus on a small subset of underserved platforms whose data are not getting onto the GTS. Tide Gauges have been identified as such a platform. They are ideal in that they are a relatively well-known platform, with relatively well-known data types. Discussions with the various networks during the OCG meeting may well help to clarify which platforms should be considered in this expanded effort.

Data Integration

OCG will continue to work with established and emerging networks to integrate their delayed mode data into ERDDAP services, as specified by the JCOMM Data Management strategy.

OCG will explore the effectiveness of making real time data available through the cloud and cloud providers. Though the immediate value of this is murky, one potential strength could be the ability for cloud providers, such as Google’s BigQuery, to enable machine learning processes on the real time data, in the context of other data holdings that exist in BigQuery and find interesting and societal connections. This would not necessarily be a benefit for the specific community that collected the data, but could benefit other communities that may not otherwise have access to the ocean data.

1. From the ToR defined for the OCG Vice-chair for Data and Information position [↑](#footnote-ref-1)