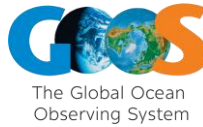


CLIVAR IORP/POGO REGIONAL TRAINING WORKSHOP ON OBSERVING  
THE COASTAL AND MARGINAL SEAS IN THE WESTERN INDIAN OCEAN  
JUNE 2022



# Data Best Practices to Support Quality Assurance and Open Science tools

*Kevin O'Brien - OCG Vice-chair for Data and Information*  
*Mathew Biddle - Data Management Analyst, US IOOS*

# GOOS OBSERVATIONS COORDINATION GROUP (OCG)

The Observation Coordination Group (OCG) works to **efficiently operate**, **maintain**, **coordinate** and **integrate** a comprehensive *in-situ* global ocean observing system

Among the FOCI for the GOOS OCG are:

- Data Management
- Standards and Best Practices

Current Data Management Activities:

- Mapping data/metadata flows for the Global Ocean Networks
- Developing a Data Strategy Implementation Plan to support WMO and IODE Data Policies/Services, FAIR data principles and improved interoperability of data
- Improve data links to Bio Eco communities



## OCG data and metadata: **Goals**

- FAIR compliance of metadata, data and data services across all OCG networks
  - Federated end point of distributed services for OCG data
- Quality data are available in near real time from the GTS and/or other data access services
- Data/Metadata are discoverable and harvestable
- Data/Metadata are available through identified global repositories
- Data are fully documented and required metadata is available through OceanOPS
  - OceanOPS services will connect to external catalogs (WMO, IODE, GOOS, UN Decade)
- Data/Metadata are properly archived and citable

# OCG data and metadata: **Data Mapping**

## Mapping the Network Data flows

- Mapped near-real time, delayed mode and metadata flows for the global in situ networks

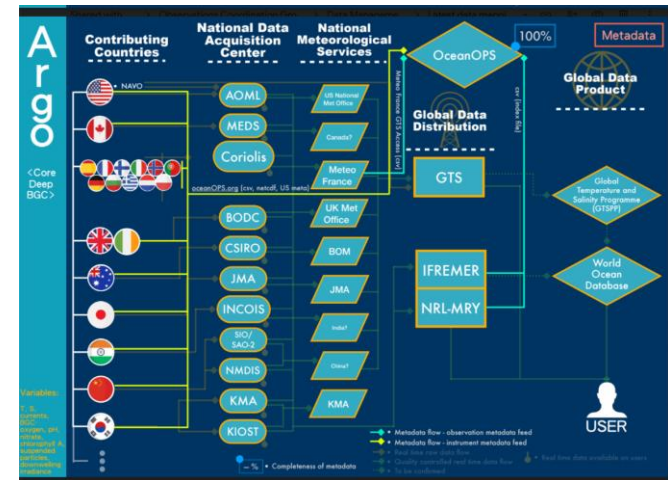
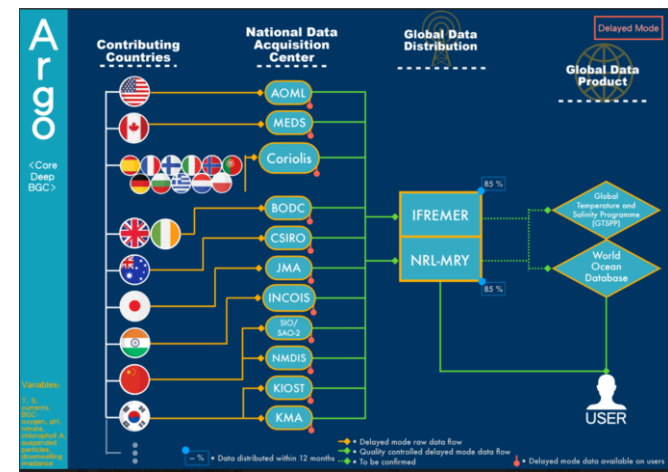
**Goal:** Identify gaps and opportunities to move OCG data and metadata services towards FAIR compliance

**First iteration** of OCG Data Mappings **complete\***

Next steps:

- **Validate** with data teams from each network
- Work with IODE, WMO to **extend** data mappings
- Use Data Mappings to develop OCG data **recommendations** and **implementation strategy**

\* will be available from new OCG pages on GOOS site soon



## OCG data and metadata: Recommendations/Best Practices

Real Time Data	Metadata	Delayed Mode	Best Practices
Real Time exchange via the WMO GTS in the approved BUFR formats	Defined uniform metadata content across platforms and/or networks	Identified Global Repository	Best Practices documenting data infrastructure and workflows
If possible, Non-GTS real time data should be available via interoperable services (ERDDAP)	Implement metadata transmission to metadata repository via m2m services	Preferred data products available through ERDDAP services	
	Where possible, metadata should be based upon standard vocabularies (CF, DarwinCore, etc)	Additional (Sensor, provenance) metadata available through global repository	Raw data, delayed mode data and data products should be archived and have DOI assigned for citation and reproducibility
		NetCDF preferred file format, though ERDDAP services can help fill that gap	

## OCG data and metadata: Recommendations/Best Practices

Real Time Data	Metadata	Delayed Mode	Best Practices
<p>Real Time exchange via the WMO GTS in the approved BUFR formats</p>	<p>Defined uniform metadata content across platforms and/or networks</p>	<p>Identified Global Repository</p> <p>Preferred data products available through <b>ERDDAP</b> services</p>	<p>Best Practices documenting data infrastructure and workflows</p>
<p>If possible, Non-GTS real time data should be available via interoperable services (<b>ERDDAP</b>)</p>	<p>Implement metadata transmission to metadata repository via m2m services</p>	<p>Additional (Sensor, provenance) metadata available through global repository</p>	<p>Raw data, delayed mode data and data products should be archived and have DOI assigned for citation and reproducibility</p>
	<p>Where possible, metadata should be based upon standard vocabularies (CF, DarwinCore, etc)</p>	<p>NetCDF preferred file format, though <b>ERDDAP</b> services can help fill that gap</p>	

# What is ERDDAP?

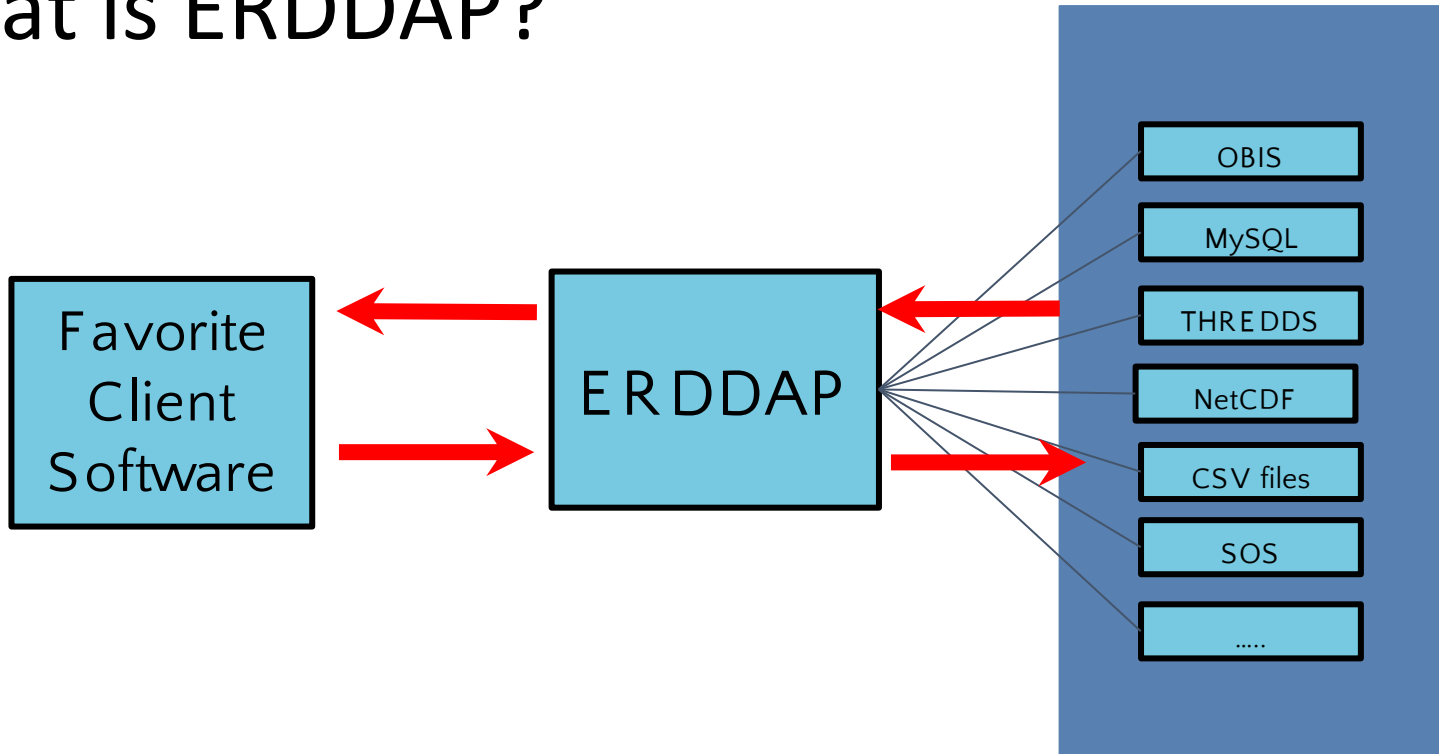
**ERDDAP is a data broker designed to help:**

- **Data providers make it easy to share data and metadata to humans and machines in many standard formats**
- **Data users to get data easily into their favorite client software without reformatting data**

**ERDDAP is open source software and freely available**



# What is ERDDAP?





# BENEFITS OF ERDDAP

- Open Source, available in github, community supported
- Much lower technical barrier – Dataset configuration, not coding!
- Supports Machine-2-Machine and Human connections
- Acts as a very effective on-the-fly data format translator - Can help with those struggling with formats such as CF - NetCDF
- ERDDAP services can be federated to support distributive data networks
- By default, supports modern data discovery standards – data provider doesn't need to do anything special
- Can provide a host of server-side functions (binning, averaging, etc) to reduce data transfer loads
- Can be used to ingest data directly! (sensor → ERDDAP)
- Provides an easy path to FAIR compliant metadata, data and services

So many more.....+ Data capacity development potential

# A FAIR-based Tool for Improving Global Data Sharing

Some ERDDAP references:

- [ERDDAP Introduction: A GOOS Webinar](#)
- [ERDDAP home page](#): Home page for documentation and downloads
- [Awesome ERDDAP](#): A curated list of ERDDAP projects and Deployments
- [ERDDAP.com](#): A Federated ERDDAP search portal
- [ERDDAPY](#): A python module for working with ERDDAP servers
- [Rerddap](#): A general purpose R client for working with ERDDAP servers
- [ERDDAP Docker Image](#): Use docker containers to install and setup ERDDAP
- [US IOOS ERDDAP configurations](#): Example ERDDAP configuration for US IOOS profiles
- [ERDDAP Github](#): ERDDAP code repository

NEXT UP:

IOOS QC and Open Science tool demonstration - Matt Biddle

