OceanGliders
Data Management

D. Hayes , V. Turpin, and the OGDMTT

The Data Management Task Team (OGDMTT)

Background

- Objective
- History
- Who we are

Progress

- Events
- Organization (TC, WG)
- Membership
- Documentation

Plans

- Continue above with more community involvement
  - Feedback from data users, glider operators, developers, companies
  - Contributions from experts
  - Financial support from agencies
  - Concrete milestones and regular meetings, yearly face-face

OGDMTT Background

Objective

- Promote and coordinate collaboration and sharing of data and tools internationally
  - Allow scientific community to find, use and re-use data reliably
  - Quantify the amount, quality, and impact of glider data for socially-relevant issues
  - Findable, Accessible, Interoperable, and Reuseable (FAIR principle)

History

- One of TT when OG was founded
- Only cross-cutting TT
- Existing efforts leveraged: EGO/GROOM, EuroGOOS TT, EMODnet Physics IOOS, IMOS

Name
Antonio Novellino
Marco Alba
Paolo D'Angelo
Justin Buck
Matthew Palmer
Mark Hebden
Emma Slater
Alvaro Lorenzo Lopez
Juan Gabriel Fernández Pineda
Miguel Charcos Llorens
Kevin O'Brien
Orens de Fommervault
Elizabeth Creed
Eric DeLory
Tania Morales
Melany Belzile
Guilherme Pimenta Castelao
Kai Salm

Who we are

- Oceanographers, data scientists, engineers
- Open to anyone, growing

OGDMTT Progress

Events

- First dedicated data meeting, Sep. 2018, Genova
  - 65 attendees
  - Principle of single data format agreed
  - Action plan in place
- Side Meeting + Breakout, New Brunswick, 27 people

OGDMTT Progress

Organization

- Technical Coordinator now hired at JCOMM OPS (Victor)
    - Goal is to help implement and track the status of the OG Network
    - Metadata are key: who did what and when, also for planning
- Main group plus topics in **Working** Groups
- Collaboration platforms and Meetings (Slack, Skype/Zoom)

Membership

- Organizing into topics to plan and achieve concrete progress
- Call to join TT and 2 WG sent out to ~30 known stakeholders (Feb 2019)
    - Currently 18 in TT, 7-8 each of 3 WG
    - WG: Format requirements (metadata and data that must be included in NC and how)
    - WG: QC (what is being done in RT and how documented, do we need to agree? DM?)
    - WG: GTS (BUFR development for forecasting systems)
    - KPI and Best Practices run throughout

OGDMTT Progress

WG OG1.0

- Harmonization of metadata and simple set of variables needed by end of year
- Could be a mapping implemented via Errdap or converters
- Possibility of redundant GDAC server?

EMODNet	Antonio Novellino
	Marco Alba
	Paolo D'Angelo
NOC	Justin Buck
	Emma Slater
	Mark Hebden
NOAA/JCOMM OCG	Kevin O'Brien
JCOMM	Victor
ANFOG-IMOS	
IFREMER	

WG QC

- Who is doing what for RT
- Document and share practices and tools
- Discuss: required OceanGliders QC?

NOC	Matthew Palmer
	Justin Buck
	Emma Slater
	Mark Hebden
SOCIB	Inmaculada Ruíz Parrado
	Cristian Muñoz Más
IFREMER	
ANFOG	
IOOS	Derrick Snowden team?
UC-San Diego	Guilherme Pimenta Castelao
University of Tallinn	Kai Salm

WG BUFR

- GTS is a separate data flow for forecasting centers that needs a glider-specific file
- WMO number is link that allows metadata tracking, otherwise no tracking
- List of forecasting variables and example NC files at present to be provided to WG to start the glider-BUFR template approval process

BODC	Justin Buck
US IOOS	Derrick Snowden
NOAA/JCOMM OCG	Kevin O'Brien
UK Met Office	Jon Turton
JCOM	Dave Berry

OGDMTT Plans

KPI and Best Practices

- Central themes TT will constantly develop
  - By end of year a small set of achievable KPI: internal and external
  - In cooperation with the other TT and for DM specifically
    - Latency, coverage in space, time, parameter, deployment planning
    - Community input by end of summer
  - Best practices identified along the way

Documentation

- Collect existing, generate as needed
  - Best practices
  - Specification documents

Continue above with more community involvement via TT or WG-Please Join!

- Feedback from data users, glider operators, developers
- Contributions from experts via WG
- Annual Data Management Meetings
- Financial support from agencies

Background: US Underwater Glider Workshop Jan 18-19, 2017

1. List requirements and/or opportunities

a. Managing metadata, calibration coefficients, etc.
b. Ability to integrate different datasets
c. All data available in same format (automation of conversation to netcdf)
d. CF compliant data at current standards
e. CF data need to be converted from cdf to profiles
f. QAQC– including agency requirements for QAPPs
g. Archive raw data (time series, delayed mode, real-time)
h. Create a group of representatives to figure out current state of data management

2. What are some of the challenges in achieving these opportunities?

a. Entry into system for new users/learning curve for building and querying netcdf files
b. Lack of resources for small shops
c. Reliance on third party (ERDAP) adds layer of complexity
d. Inconsistency in metadata between groups
e. Time/money for personnel at each group to prep data
f. Inconsistent QAQC
g. Lack of automation to get files into needed formats (not provided by manufacturer)
h. All gliders output data in different formats
i. Standards change
j. High configurability of gliders means there are sensor data with no standard name, and regularly new types of data
k. Archiving raw data- ownership
l. Coordinating communication between the range of user groups
m. Central location for housing information

3. Provide recommendations/next steps

a. Establish minimum metadata, etc. standard (1-3 year)
b. Coordinate with EGO and IMOS to develop a unified system (a la ARGO) or improve compatibility – bring together representatives from all 3 groups to work on this (EGO building a team), can also improve QC; include manufacturers in these discussions (1-3 year)
c. Centralize data management and formatting within the regions (1-3, after initial meeting)
d. IOOS should develop set recommendations for manufacturers (1-3, after initial meeting)
e. Recommendation to manufacturers to provide capability to convert data into netcf when they get to shore (Establish minimum requirements to be implemented now, and wish list of further capabilities
f. Training at the universities (ongoing, coordinate with pilot training programs)
g. Quantify value of standardized formats – who's using the data? How? Etc.
h. DOI for each deployment to track use of data (developing this at EGO and OOI, publications are starting to require this)
i. Include QAQC plan into funding requests
j. Discussion on archiving raw data and relation to ownership
k. Data management team –road map, develop standards, implementation, training
l. Side meetings at major conferences –AGU, Oceans Sciences, etc.

Background: **EuroGOOS/EMODnet Data Flow** 18-19 Sep., 2018

https://www.ego-network.org/dokuwiki/doku.php?id=public:egodmmeeting:september2018

Action Plan

- The creation of a single global glider format, OceanGliders1.0
- The definition of a global objective for glider data management and the formation of KPI
- QC procedures must be defined, harmonized and documented.
- An international team will be created build to collect of OceanGliders best practices
- A first version of updated community tools and reference documents
- Monthly meetings and physical (EGO8), and dedicated data management (??)

Background: **EuroGOOS/EMODnet Data Flow** 18-19 Sep., 2018

Take Aways

 The harmonization of formats towards a single OceanGliders data format was decided as a priority;
 An open governance structure and format development/release process to harmonized format, implifying data submission, data interoperability and future format development should be set up in parallel;
 It has be stated that there is a need for Key Performance Indicators (KPIs) to assess the global glider network. Some KPIs can be calculated from the data management system, covering the following aspects such as : planning, real time data delivery, delayed mode data delivery and variables;
 Governing structures exist (EuroGOOS GTT, OceanGliders Steering Team, JCOMM OCG, CMEMS-INSTAC) and should be solicited for guidance, to monitor progress made by our community and used to endorse strategical choices
 Documents are needed for cooperative sharing and development (data management plan, including real time and delayed mode quality control manuals);
 Community tools for data management should be in line with the current format, easily accessible and well documented to allow easy access to the data management process;
 Data management activities should be accounted for in budget estimates as the recommended practices require significant human resources;
 A glider technical coordinator is absolutely necessary to ensure accurate metadata are collected and stored;
 Regular meetings of international data management team are mandatory (annual). Outcomes of such meetings must be reported and discussed during OceanGliders Steering Team (annual) meetings and vice-versa
 The role of JCOMMOPS is central and should be consulted regularly until a glider technical coordinator is hired;
 A long term sustained collaboration between EMODnet Physics and JCOMMOPS is important to facilitate and speed up the process and provide the EuroGOOS GTT the framework to develop and achieve the planned actions,
 The data management community will be asked to contribute to the production of the delayed mode, best practices reference documents. The community should be strongly involved in the EuroGOOS task team.
 Criteria that define an OceanGliders deployment should be clarified e.g. open data supplied in files where format is compliant so only OceanGliders program deployments are notified to OceanGliders program and deployments that do not meet the criteria do not negatively impact on metrics. These criteria will need updating as other elements of the data system such as delayed mode quality control are formalized.


Thank
you!

OceanGliders

OceanGliders