Global Glider Data Management Strategy and Implementation Plan
OceanGlider Data Management Task Team (OGDMTT)
Co-chairs: Daniel Hayes, Victor Turpin

Contents

1. Objectives

The OceanGlider Data Management Task Team (OGDMTT) is proposed here in order to develop a strategy and implementation plan for international coordination. Many important issues are important to be solved to allow true interoperability of ocean glider data and metadata regardless of the location of the data center or model of glider used. Examples include merging near real time (NRT) and delayed mode (DM) data streams, harmonizing quality assurance and control (QA and QC) in both cases, finding ways to efficiently handle technical and engineering data and meta data, identifying biogeochemical data handling, standardization of formats and protocols, among many others.

2. Expected Impacts and Benefits

Coordination: The OceanGliders Steering Committee reflects the larger glider community when they expressed their desire to establish the DMTT, since it is a cross-cutting theme, that could become a bottleneck for the other task teams and the community as a whole. Significant efforts have already been made at a number of major national and regional data centers, such as Canadian Glider Network-Ocean Networks Canada (CGN-ONC), Australian National Facility for Ocean Gliders (ANFOG), US National Glider Network (IOOS-NGN) and Everyone's Gliding Observatory (EGO) and others. In order to streamline and "globalize" glider data, it is important to provide a means for any data management stakeholder to exchange complementary skills, knowledge and tools, while avoiding duplicating effort.

Visibility: by joining forces, we increase our visibility, contribution, and influence in the international arena of the Global Ocean Observing System (GOOS) and maintain high level ties with the organizations and efforts to coordinate all kinds of earth observation data at the international level.

Harmonization: as mentioned above, harmonization of QA/QC and the definition of a common metadata basis are required for future glider data/metadata interoperability. Already there are at least three major glider data formats, and the OceanGliders DMTT platform provides an opportunity produce a detailed analysis of each of the primary file formats from the glider vendors and the DMAC groups and possibly design a format conversion tool between them. This would make it possible to reconcile data and metadata between the major data repositories (either towards a merged format/consistent fields or a mapping from one to another). At the data provider and principle investigator level, harmonizing citation practices will ensure credit is given where credit is due when it comes to data collection, processing, and provision.Technical aspects of harmonization will be addressed, such as:

- increased use of recognized controlled vocabularies (e.g., some of the NERC 2.0 ones),
- introduction of globally unique identifiers (in addition to the wmo-id, ORCIDs for people, DataCite dois, mission identifiers, etc)
- Web services for better linked metadata/data might also be a useful direction - not just final data products, but mission metadata, tracklines, etc.

3. Members

Our strategy is to contact every possible expert in data management who has an interest in glider data and metadata, both scientific and technical and invite them to take part in this exchange. Everyone who wants can have input into the plan developed now, and moving forward after this task team is formally closed. (Each task team is expected to last two years.)

Up to now, the people interested in glider data management are registered in the og-dm@jcommops.org mailing list and can contribute to the slack discussion on demand through the mailing list or directly to dhayes@ucy.ac.cy od vturpin@jcommops.org.

4. Mode of Operation

The OGDMTT will meet twice per year witha teleconference and/or face to face meeting. A shared cloud space, possibly at the EGO web site, where users can register and upload/download files, will be provided to maintain reports and member lists.

5. Interactions with others

Regular contact with the data management experts around the globe must be maintained in order to take advantage of harmonization and visibility opportunities as they come. To do this a list of meetings or conferences of each TT member (e.g. a google calendar) will be maintained, which will allow ad hoc meetings to occur even if there is not a devoted session or meeting planned. It is important to build a sense of collegiality and cooperation since up to now, most major data centers have worked independently when it comes to glider data management.

6. Milestones and proposed action items

The first milestone obviously, is to jointly prepare and submit this proposal. Contacting the interested people and collecting their feedback took a significant amount of effort. We will provide an update progress report by end of year 1, present at EGO meeting or similar, with a final report end of year 2. This will be approved and posted.

Specific Tasks:

- Identify contact people, contribution, domain (in progress). Spreadsheet to be updated accordingly (every 3 months).

- Detailed analysis of each of the primary file formats from the glider vendors and the DMAC groups (report, after 6 months)

- Identify workflows at each DMAC group (1-2 pages each, after 12 months)

- Identify possible ways to reconcile data formats and workflows using standards when possible (5-10 pages, after 18 months)

- Approve a technical strategy and first implementation plan to begin in the future, as resources allow, in which the most agreeable methods of harmonization can be achieved.

- Approve a strategy for the future organization of the OGDMTT that will carry its interests and objectives beyond the TT life cycle into a permanent platform for exchange and collaboration.

References

US IOOS National Glider Data Assembly Center.: https://github.com/ioos/ioosngdac and documentation on the design, file formats, data submission is here:
https://github.com/ioos/ioosngdac/wiki

EGO format: http://www.seanoe.org/data/00343/45402/