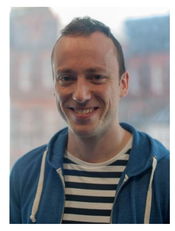
Progress towards the establishment of an EGO Data Assembly Centre (DAC) for UK glider data

British Oceanographic Data Centre
NATURAL ENVIRONMENT RESEARCH COUNCIL

Mark Hebden (mahe@bodc.ac.uk), Justin Buck (juck@bodc.ac.uk), Louise Darroch, Thomas Gardner and Alexandra Kokkinaki

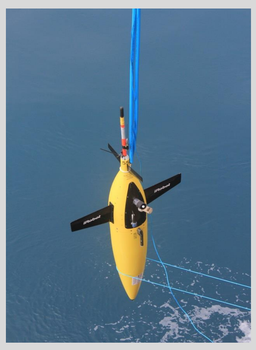


Overview

Our Goals

- To **safeguard** ocean glider data and metadata on behalf of the UK glider community.
- To **standardise** and **disseminate** quality-assured, open-access UK glider datasets to a wide range of stakeholders, in a streamlined and timely fashion.

"Capture once, use many times"



Abstract

BODC is committed to exposing UK glider data observations through the Everyone's Gliding Observatories (EGO) initiative. Efforts to date have focused on safeguard and dissemination of interoperable data for long term re-use. BODC's current work is presented alongside our future aspirations, which include championing an 'open data' culture to maximise impact of UK glider capabilities.

Background

BODC is the designated Data Assembly Centre (DAC) for UK glider measurements. The UK glider fleet is growing rapidly and associated sensor technologies are emerging and maturing at a similar pace. Development of robust infrastructure and processes at the DAC is necessary to ensure effective archive and long term exchange of quality glider data to end users.

The EGO framework provides glider DACs with standards to govern glider data appropriately, as well as a conduit to expose datasets to a wide international audience.

DAC Workflow (red = work in progress)

Safeguard → **Standardise** → **Disseminate**

Acquisition

- Engaging with glider operators to ensure timely exchange of platform and deployment metadata/data.
- Flexible secure methods (e.g. SFTP and rsync).

Archive

- Near real time (NRT), post-recovery and delayed mode (DM) versions.
- Multiple copies stored, including offsite. 'Failover' strategies.

Project distribution

- Secure file transfer service to provide single conduit to data/metadata for project stakeholders.

BODC 'NRTDS'

- A generic **N**ear **R**eal **T**ime **D**elivery **S**ystem (NRTDS) to handle various platforms, including gliders. Work commenced 2016.
- JSON metadata input, EGO V1.1/1.2 data exchange format output.
- Legacy SAMS Seaglider and MASSMO project datasets are currently providing the test case for BODC.
- BODC are expanding the SeaDataNet vocabularies for technical parameters through this work.

Governance

- BODC involvement with international projects and initiatives to ensure we contribute (and adhere) to current standards and protocols for glider data.

Routine DAC data pushes

- NRT delivery to the UK Met Office and Global Telecommunication System (GTS). Currently in WMO FM-64 TESAC format.
- Delivery to the EGO Global Data Assembly Centre (GDAC), Coriolis.

Tools

BODC are working on using and contributing to common tools to deliver data.

- EGO files are produced using software developed collaboratively with Ifremer
- US NOAA produced ERDDAP to visualise, aggregate and subset data
- 52North to provide OGC SWE compliance.

The tools will first be used to deliver SenseOCEAN project data.

Data Interoperability

Data are to be delivered using common standards including:

- W3C linked data
- OGC Sensor Web Enablement (SWE)

The SenseOCEAN and BRIDGES projects enabled BODC to collaborate with international partners to produce the "Marine SWE profile" - a harmonised implementation of SWE ensuring interoperability between participating data groups by including semantic mark up within metadata.

Associated web services will allow us to expose datasets via the web in an interoperable manner.

DAC Next Steps

'Oceanids'

The four year long 'Oceanids' project kicks off in September 2016 and aims to establish a seamless data flow from sensors on autonomous platforms through to data users.

BODC will be working closely with the UK's National Marine Facilities Sea Systems (NMFSS) as part of the project, with the resulting infrastructure benefiting the wider UK glider community.

This work will be key to achieving open data enabling practices…

Open Data

There is an international move towards open data across scientific and humanity disciplines. The recently published ICSU document on Open Data in a Big Data world, an international accord summarises the enabling practices as:

- Citation and provenance
- Interoperability
- Non-restrictive reuse
- Linkability

Many of these practices are encompassed in current BODC development and the oceanographic community moving towards non-restrictive reuse. A major precedent was set by the Argo project 18 years ago with open data from the outset. NERC introduced the open government licence once data have passed a two year project restriction. Further to this SAMS have granted open access to glider datasets, including those collected on the Extended Ellett Line.

DAC Priorities

- Further develop and streamline the workflow for UK glider data.
- Push EGO-formatted SAMS data to GDAC.
- Engage further with UK glider users and forge closer working relationships, particularly with BAS.
- Switch to EGO format delivery to UK Met Office.

Glider Workshop - Foresight Centre, Liverpool. 18 April 2013