

Future data delivery from autonomous platforms to data users; Applying OGC Sensor Web Enablement and Linked Data standards

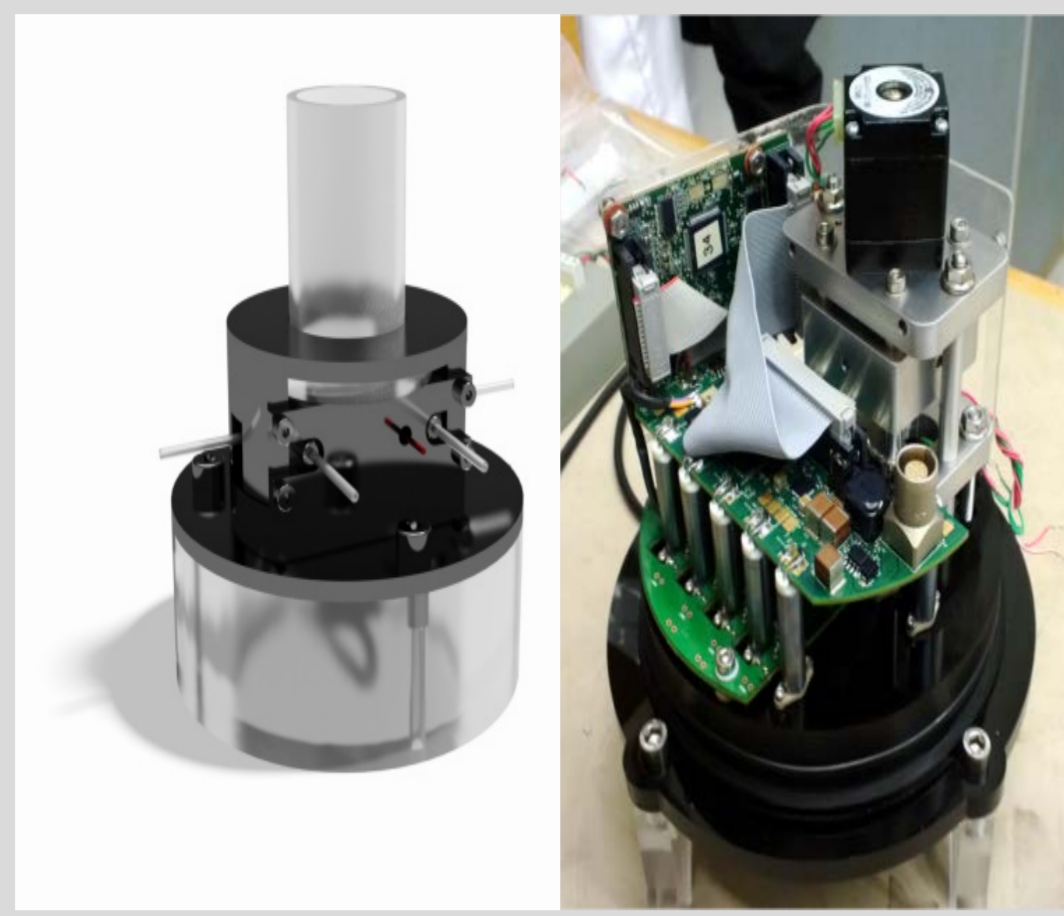


Justin Buck (juck@bodc.ac.uk)
Mark Hebden (mahe@bodc.ac.uk)



Glider Workshop - Foresight Centre, Liverpool. 18 April 2013

Data sources



Sensor metadata

As part of the SenseOCEAN project BODC are working on the standardisation of sensor metadata enabling 'plug and play' sensor integration. Sensors will either transmit their metadata in SensorML directly via the glider or send a unique link to a SensorML record at the data centre (minimising transmission costs and bandwidth).

Abstract

Oceanic gliders are now an established tool for observing the ocean for both operational and research purposes. Glider capabilities are continuing to rapidly evolve and data volumes/complexity are consequently increasing.

BODC have recently commenced a new batch of projects that will evolve the data management and transfer from a process that requires significant manual intervention to an increasingly automated operational setup. The primary method for doing this is to apply standards to data all the way from the sensors through to data delivery.

This poster shows the current capabilities of BODC (**blue titles**) and on-going development/newly funded activities from recent proposal successes (**orange titles**).



Platform metadata

On the BRIDGES project the use of OGC Sensor Planning Service, metadata delivery in OGC SensorML and data delivery on OGC Observations & Measurements formats is being investigated. If applied across multiple platform types it will reduce the number of source data formats.

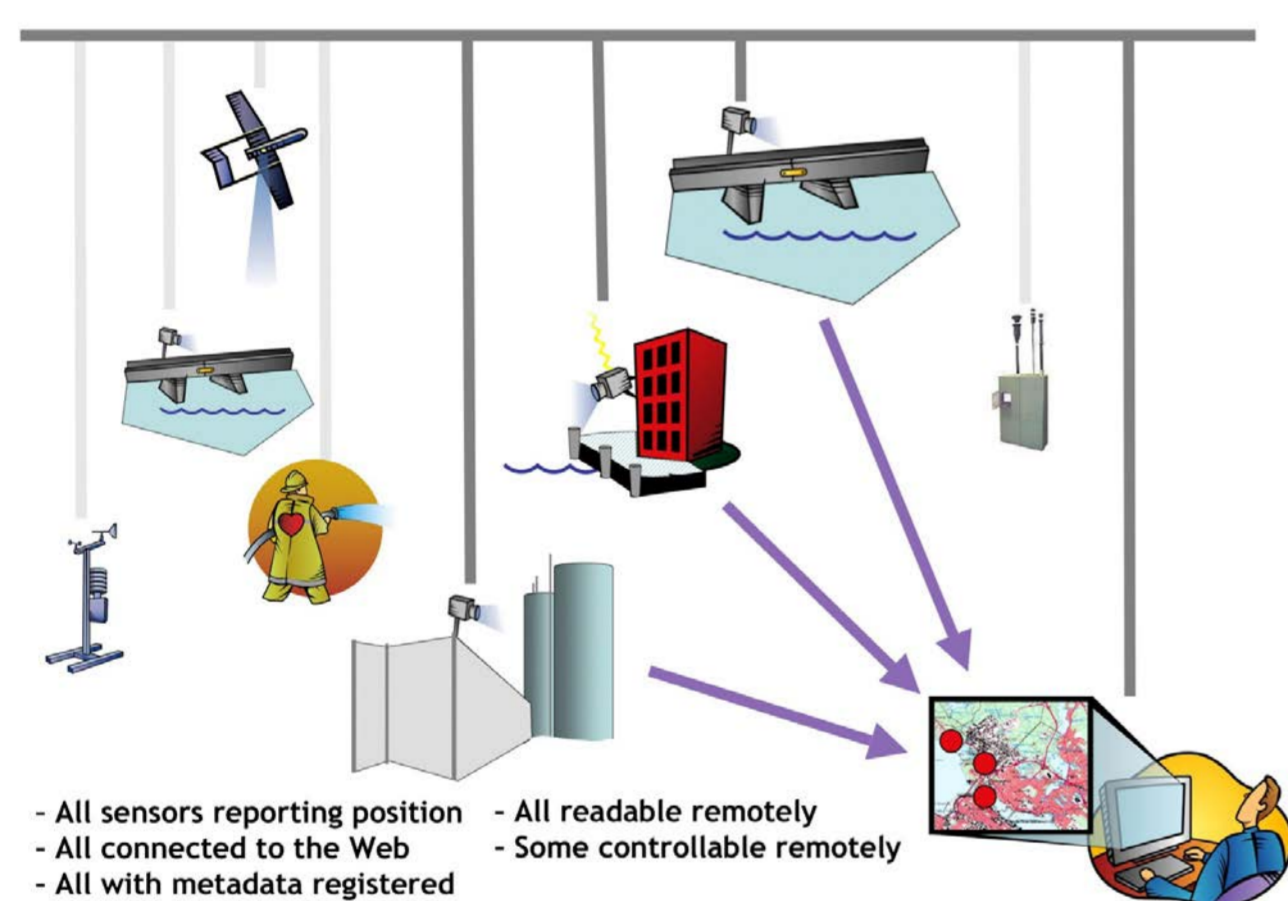
Data assembly centre

Sensor Web Enablement



The OGC's Sensor Web Enablement (SWE) standards enable developers to make all types of sensors, transducers and sensor data repositories discoverable, accessible and useable via the Web.

BODC are part of four EU projects that will implement OGC SWE (SenseOCEAN, BRIDGES, AtlantOS, ODIP). These tools will make data exchange and usage more efficient for data originators, data centres, and end users.



A simple schematic showing the principles of OGC SWE for connecting data from <http://www.opengeospatial.org/ogc/markets-technologies/swe>

Current BODC capability

The BODC is the designated Data Assembly Centre (DAC) for UK glider data.

Data are acquired by BODC in near real time using various methods (including ftp, wget and rsync) and securely archived automatically.

The archival process involves storage of multiple backups of the data, including offsite copies.

Code, developed in collaboration with the international glider community, is used to rationalise the various glider data formats into a single 'data exchange format', the Everyone's Gliding Observatories (EGO) NetCDF. The EGO format pulls together the glider data and associated metadata into a single self-contained file and promotes data interoperability. EGO NetCDF conversion is carried out by BODC.

Reformatted data can be supplied back to originator (via password-protected ftp) as the basis of further post-processing of glider data in a delayed mode capacity.

Linked data

W3C Linked Data is a means to publish structured linked data via the web. BODC's goal is to publish 'five star' linked data.

- ★ Data available on-line, through the World Wide Web
- ★★ Data available in a machine readable, structured format: e.g. a spreadsheet, not an image of a table
- ★★★ Data are supplied in a non-proprietary format, e.g. Comma Separated Values, instead of MS Excel
- ★★★★ Data are supplied in using open standards from the World Wide Web Commission
- ★★★★★ Data are published with links to other datasets in order to provide context

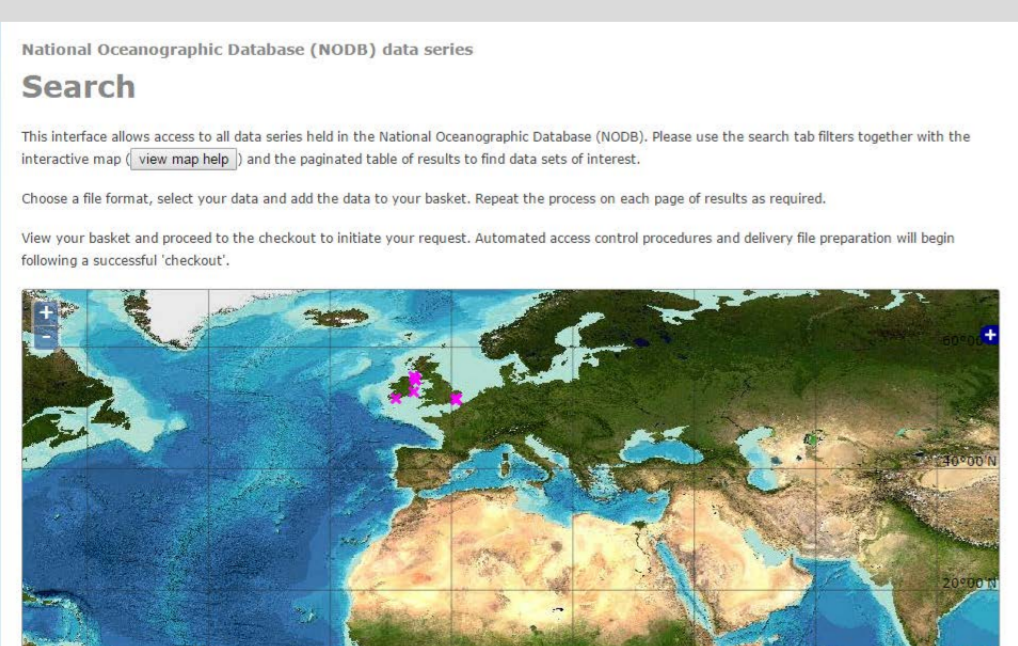


An example of such linkages is the BODC vocabulary server which links to external sources to fully describe vocabularies. The vocabulary server is then linked to by external websites to be included seamlessly within their web content (schematic courtesy of Adam Leadbetter, Marine Institute, Galway, formerly BODC). <http://www.w3.org/standards/semanticweb/data>

Data delivery

Web services

Glider data are to be exposed to the web via webservers (one OGC Sensor Observation Service and one W3C Linked Data) that enable 'machine to machine' access to data. This will make the data automatically available to data assimilators.



Web

Once OGC SOS and Linked Data are operational, glider data will be made available via BODC's web pages at <http://www.bodc.ac.uk>.

Operational data delivery

BODC can repackage near real time glider observations into the WMO FM-64 TESAC format for submission to the UK Met Office. TESAC messages contain low resolution profiles of temperature and salinity. The Met Office ingest glider observations into their operational database, MetDB, from where they are available for assimilation into the FOAM (Forecast Ocean Assimilation Model) system. Data are also routed to the Global Telecommunication System (GTS), a means of sharing the data with Met Offices worldwide in near real time.

Future data delivery

BODC are preparing to supply the UK Met Office with real time EGO NetCDF glider data (in near real time) in future for in-house conversion to the BUFR format. Full data resolution can be achieved with BUFR and it also paves the way for widening the range of data variables that can be shared.

Data restrictions

BODC are fully aware of and respect the needs of data providers to restrict access to datasets until they have been exploited by projects that fund glider deployments. This is according to NERC data policy. Prototyping of new systems described here will use data that have been designated as unrestricted by data originators.