Interoperable Management of Glider Sensor Data

Matthes Rieke
52°North Initiative for Geospatial Open Source Software GmbH
Münster, Germany
m.rieke@52north.org

Simon Jirka
52°North Initiative for Geospatial Open Source Software GmbH
Münster, Germany
jirka@52north.org

Daniel Hayes
Cyprus Subsea Consulting & Services Ltd.
Nicosia, Cyprus
hayesdan@cyprus-subsea.com

Justin H. Buck
British Oceanographic Data Centre
Liverpool, United Kingdom
juck@bodc.ac.uk

Ehsan Abdi
Cyprus Subsea Consulting & Services Ltd.
Nicosia, Cyprus
e.abdi@cyprus-subsea.com

Abstract

The management of observation data gathered by autonomous underwater vehicles such as gliders is currently lacking an interoperability concept for the discovery, access and interpretation of data. While EGO NetCDF is a well-established format for data gathered for single missions, it does not comprise a complete data management system and therefore cannot be expected to fully address the aforementioned features. In this article we present an interoperable system design based on the concepts of the Sensor Web as defined by the Open Geospatial Consortium. The usage of the Sensor Observation Service as a discovery and access interface in combination with ISO/OGC Observations & Measurements for encoding glider missions is illustrated. We focus on the specifics that need to be addressed to achieve the overall goal of data interoperability within the ocean glider domain. In addition, a concept for semantic interoperability of glider observation data using vocabularies is introduced.

Keywords: Glider, Autonomous underwater vehicles, Sensor Web, Open Geospatial Consortium, Ontologies

1 Introduction

Managing ocean observation data gathered by autonomous underwater vehicles (AUVs) has always been a challenge. Several vendors compete in the market, all providing their own software solutions. Data access and subsequent processing therefore highly depend on the platform used. To address the data management issues for long-range AUVs known as gliders, oceanographers and engineers of the Everyone's Gliding Observatories (EGO) network have initiated streamlining activities such as the definition of the EGO NetCDF format. While the use of this format has grown in recent years, there is still a set of obstacles that prevent glider data from being exploited efficiently. The ability to discover and identify relevant glider data for a specific analysis is still limited as is the access to it. In addition, the processing and interpretation can be very specific to a mission, a platform or the sensors used to gather the observation data.

In this work, we address these obstacles with a system design for managing glider observation data based on concepts of the Sensor Web as defined by the Open Geospatial Consortium (OGC). We illustrate the use of OGC Sensor Observation Service (SOS) in combination with SensorML and Observation & Measurements (O&M) and how they help to overcome the aforementioned interoperability issues. Here, vocabularies play an important role as they provide means to semantically interpret data from different platforms or missions in a common way.

The designed system architecture has been developed within the BRIDGES Horizon 2020 project. BRIDGES does not only address data management aspects but aims at developing two new types of glider platforms: 1) the Deep Explorer with a maximum diving depth of 2400 meters, 2) the UltraDeep Explorer which will be able to cover depth of up to 5000 meters. The capability to dive into these regions enables a set of new applications for gliders, in particular various scientific and industry-driven use cases. As the BRIDGES project is ongoing, this article presents work in progress and sheds light on the upcoming challenges within the development of interoperable glider data management. The designed interoperable data management system can be applied to existing glider platforms as well as to the newly developed AUVs.

2 Related Work

2.1 Sensor Web

The term Sensor Web as used in this work refers to the integration of sensors and their data into web-based infrastructures (Bröring et al., 2011). This comprises for example the sharing of sensor observation data via the web or the control of online sensors. In order to ensure interoperability within such Sensor Web infrastructures, the Open Geospatial Consortium (OGC) has developed a range of standards which form the so-called OGC Sensor Web Enablement (SWE) framework (Botts et al., 2006). This framework comprises standards addressing both the specification of web service interfaces as well as the definition of data models and encodings.

The SWE web service interface standards most important for this work are the OGC Sensor Observation Service (SOS) (Bröring et al., 2012) as well as the OGC SensorThings API (STA) (Liang et al., 2016). Both of these specifications enable the access to observation data, but with different foci. While the STA is especially optimized for IoT applications with lightweight communication mechanisms (i.e. REST/JSON as well as MQTT) the SOS standard is typically used to transfer XML encoded observation data that follow schema specifications which are adjusted to the needs of official data reporting flows or domain-specific data models.

Complementary to these interface specifications, the OGC SWE framework also offers a range of data models and encodings. The modelling and encoding of observation data is covered by the Observations and Measurement (O&M) standard. This standard is composed of two separate specifications: an ISO standard defining the conceptual O&M data model (ISO, 2011) (e.g. defining the structure of observations and their different properties) as well as an OGC standard describing how O&M data shall be encoded in XML (Cox, 2011). Besides observation data, a second aspect of data models concerns the description of the sensors and measurement processes which were applied to obtain observation data (commonly referred to as observation metadata). This is especially important for discovery purposes and for ensuring the correct interpretation of observation data sets. To cover this requirement, the OGC SWE framework offers the Sensor Model Language (SensorML) specification (Botts and Robin, 2007).

Different approaches for applying Sensor Web technologies in marine sciences have been and are being actively investigated by a range of different international research projects. For example, the NeXOS project (Toma et al., 2015) investigated the development of a full plug-and-play integration chain for sensors and marine sensing platforms. The SeaDataCloud project (Schaap and Fichaut, 2017) aims at improved interoperability of ocean observation data provision within marine research data infrastructures. As a third example, the SenseOcean (http://www.senseocean.eu/) project works on an in-situ marine biogeochemical sensor system which makes use of SWE-based interoperability. There is a broad range of further relevant research projects, which cannot be presented because this would go beyond the scope of this article.

2.2 Everyone's Gliding Observatories

Everyone’s Gliding Observatories (EGO) is an international network of partners in the oceanography domain dedicated to the promotion of glider technology and its applications. Besides coordination, workshops and training, EGO also works on the definition of a NetCDF profile (EGO 2017) to streamline the storage of glider-gathered data sets. It covers data structures for navigational properties (trajectory data based on GPS and inertial sensors), global attributes (spatial and temporal coverage, producer information, etc.) and the actual measured data variables (e.g. salinity, water temperature). The system architecture presented in this article makes use of the NetCDF EGO profile, benefitting from its establishment within the glider community.

The current version of the EGO gliders User's manual defines a workflow for data management and access that takes the roles of three organizational units (Principal Investigator, Data Assembly Centre (DAC), Global DAC) into account. The goal is to ensure data quality among different glider operators and to provide transparency on the quality of data sets to users. All data is gathered centrally at Ifremer (hosting the Global DAC) and is generally made freely available.

EGO is currently in the phase of formally transitioning to OceanGliders (http://www.oceangliders.org/). As part of the transition an OceanGliders Data Management Team (OGDMT) has been formed which will progress key themes such as data formats and quality control.

2.3 Glider Platforms and Data Acquisition

Gliders from different vendors all implement specific data acquisition workflows. Therefore, the analysis of these provides the basis for developing an interoperable approach. We considered three well-established glider platforms during this analysis: Seaglider (Kongsberg 2019), Slocum (Teledyne 2019) and SEAEXPLORER (Alseamar 2019). All three systems provide the capability to transmit data gathered in one dive via IRIDIUM satellite connection. The used protocols differ greatly, and dedicated base stations located on the shore or on a supporting boat are required to interpret the data flow. As a lot of processing logic has been put into the software of the base stations, changing the internal acquisition workflow would not be realizable. The designed system architecture is set to work on the post-processed data. All three analysed glider platforms provide the post-processed data as NetCDF, having similar but not identical structures.

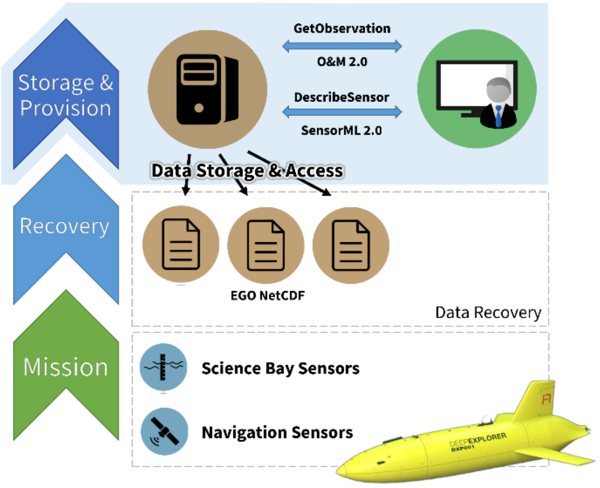
3 Challenges

During the analysis phase of this work, we have identified a set of research challenges in the field of glider data management. In particular, these are:

3.1 Interporeable Discovery and Access

How can sensor data gathered by glider platforms be discovered and accessed in an interoperable manner? The aforementioned diverse data acquisition methods and resulting data files highlight the need for a streamlined approach. We describe an approach using OGC Sensor Web technology as a possible solution for this challenge.

Figure 1: High-Level Architecture of the BRIDGES Sensor Web Approach.



3.2 Semantics

How can glider data gathered by different platforms be interpreted uniformly in a common manner? Making glider data discoverable and accessible is the first step. The meaningful processing of data from different platforms requires additional concepts such as the usage of vocabularies (e.g. for instruments, observed variables) and metadata descriptions.

3.3 Dissemination Patterns

Can acquisition patterns for glider data be designed to efficiently provide observations in near-real time? We elaborate on the possibility to use a Publish/Subscribe message exchange pattern for data dissemination.

4 Approach

In order to make the captured data discoverable and accessible in an interoperable way, we developed a system architecture based on concepts of the OGC Sensor Web. An SOS instance is responsible for storing the data of individual glider missions and provide access to these using standardized web interfaces.

As illustrated by Figure 1, the interoperable data management layer reuses the existing data acquisition and recovery mechanisms of individual glider platforms. The derived EGO NetCDF files are ingested into an SOS instance, thus allowing users to discover and access the data. The data is retrieved using the *GetObservation* method. The response is provided as an O&M 2.0 document which does not include the observation as in-line content, but refers to the original EGO NetCDF file. The benefit of this approach is that the strength of both technologies are exploited in an efficient manner:

- The usage of O&M allows the inspection of mission parameters such as the observed feature (e.g. the Baltic Sea), the observed ocean variables or the temporal extent while not interfering with the NetCDF format.
- Providing the EGO NetCDF file as a references enables further processing in already established analysis tools.

A service architecture that focuses on the discovery and provision of glider data requires a strong focus on the metadata of glider platforms, missions, as well as features of interest. An important requirement therefore is to identify the metadata required by data analysts which need to be exposed by the system. The EGO NetCDF profile covers platforms specifics such as calibration parameters for deployed sensors. The definition of observed data variables is also part of the profile. In addition to the data, users are able to retrieve metadata for gliders used in specific missions. The *DescribeSensor* method of the SOS provides these metadata encoded as a SensorML 2.0 document. It contains information such as the manufacturer serial number of the glider, operator contacts, deployment history and the installed sensor instruments.

Closely related to the provision of metadata is the handling of semantics of sensor descriptions and observation data sets in order to enable meaningful analysis. This concerns a broad range of aspects ranging from the common use of terms for referring to observed properties (i.e. avoiding different text strings that refer to the same property such as "water temperature", "water temp.", "temperature of the water", etc.) to the uniform and machine-readable description of sensor interfaces. For this purpose, the BRIDGES Sensor Web approach makes use of the NERC Vocabulary Server (https://vocab.nerc.ac.uk). This vocabulary server provides an extensive vocabulary of terms related to ocean sciences but also of important concepts that are part of the OGC Sensor Web Enablement framework. By referring to the relevant terms in O&M encoded observation data documents (i.e. for observed properties) and in SensorML-based metadata files (e.g. for describing sensor properties and capabilities, events in the history of a sensor, inputs and outputs, as well as parameters of the measurement process) a foundation for enabling semantic interoperability is achieved (Kokkinaki et al., 2016). As part of the system design a mapping from existing EGO data variables to the NERC vocabulary has been developed. Listing 1 demonstrates the usage of vocabularies within O&M encoded glider mission data. Similarly, Listing 2 illustrates how ocean variables are encoded in O&M using vocabulary definitions.

Listing 1: Glider Mission Data Encoded as O&M 2.0.

```
<om:phenomenonTime>
    <gml:TimePeriod gml:id="phenomenonTime-glider-0123">
        <gml:beginPosition>2016-01-01T00:22:00</gml:beginPosition>
        <gml:endPosition>2016-01-01T12:33:00</gml:endPosition>
    </gml:TimePeriod>
</om:phenomenonTime>
<om:resultTime>
    <gml:TimeInstant gml:id="resultTime-glider-0123">
        <gml:timePosition>2016-01-01T12:45:00</gml:timePosition>
    </gml:TimeInstant>
</om:resultTime>
<om:procedure
    xlink:href="http://www.bridges-h2020.eu/gliders/vendorA/0123"/>
<om:observedProperty xlink:href="http://www.bridges-
    h2020.eu/gliders/vendorA/0123/science"/>
<om:featureOfInterest
    xlink:href="http://vocab.nerc.ac.uk/collection/C16/current/21a/"/>
```

Subsequently, this approach can be used in different data provision, management, access, and visualisation tools. For example, metadata editors such as 52°North *smle* support the NERC Vocabulary Server so that users can insert terms from the vocabulary into sensor descriptions. Complementary to this, there are prototypical versions of the 52°North *Helgoland Sensor Web Viewer* which are capable of resolving URLs pointing to vocabulary entries in order to present users with meaningful data descriptions.

Listing 2: Variable Data Encoded as O&M 2.0.

```
<swe:field name="sea_water_temperature"
    xlink:href="http://vocab.nerc.ac.uk/collection/OG1/current/TEMP/">
    <swe:Quantity>
        <swe:label>TEMP</swe:label>
        <swe:uom code="deg"
   xlink:href="http://vocab.nerc.ac.uk/collection/P06/current/UPAA/"/>
    </swe:Quantity>
</swe:field>
<swe:field name="sea_water_electrical_conductivity"
    xlink:href="http://vocab.nerc.ac.uk/collection/OG1/current/CNDC/">
    <swe:Quantity>
        <swe:label>CNDC</swe:label>
        <!--// Siemens per meter //-->
        <swe:uom code="S/m"
   xlink:href="http://vocab.nerc.ac.uk/collection/P06/current/UECA/"/>
    </swe:Quantity>
</swe:field>
<swe:field name="sea_water_practical_salinity"
    xlink:href="http://vocab.nerc.ac.uk/collection/OG1/current/PSAL/">
    <swe:Quantity>
        <swe:label>PSAL</swe:label>
        <!--// practical salinity units //-->
        <swe:uom code="[psu]"/>
    </swe:Quantity>
</swe:field>
```

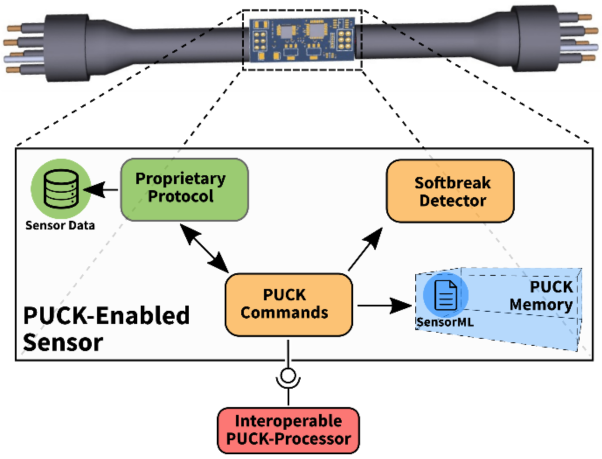
In the future the presented approach for the integration of semantics will be extended to cover further ontologies and vocabularies. While the NERC Vocabulary Server was chosen as an initial candidate due to its broad range of observed property definitions and basic Sensor Web concepts, further sources such as the Semantic Sensor Network (SSN) ontology and the Marine Metadata Interoperability Project Semantic Web Services will be considered as future work.

5 Discussion and Future Work

Within this article the current state of glider data acquisition and management has been outlined. Based on this status quo an architecture for interoperable data discovery and access has been developed. Here, the focus was laid on well-established international standards such as the OGC Sensor Web Enablement standards suite. Conceptual development and first tests have been performed to proof the technical approach. Currently, the system has been designed but not yet tested in real-world applications. However, the technological chain was proven with test data. A full paper about further experiences gathered and possible subsequent adjustments to the architecture, is planned.

In the course of the BRIDGES project possible use cases have been identified where the provision of near-real time data gathered by gliders could be beneficial. For example, an oil spill observed and detected by a glider would require immediate action whereas the manual recovery and post-processing of data would not address the time criticality in a sufficient way. The OGC Publish/Subscribe 1.0 Interface Standard (Braeckel et al. 2016) has been identified as a suitable concept that allows the push-based dissemination of geo-data. As it is a generic standard it can be applied easily to the Sensor Observation Service and thus to the system design presented in this work. An alternative approach that is planned to be realized in future work is a design based on the STA. Its MQTT binding extension is a promising candidate for push-based dissemination. In the remaining time of the BRIDGES project we will work on a prototypical design that leverages the IRIDIUM transmission capabilities to establish near-real time data provision. Here, the assessment of automated data quality control will be one of the major research challenges.

Figure 2: PUCK-Interoperability with the SMART CABLE.



Additional future work will investigate a lower level of interoperability. The PUCK Protocol Standard of the OGC enables the automated configuration and data retrieval on the sensor layer. As the majority of sensors available on the market do not implement the PUCK Standard, a generic approach has been prototypically designed and will be further extended and validated on glider platforms. The SMART CABLE, developed by partner Cyprus Subsea Consulting and Services C.S.C.S. Ltd., aims at introducing a microcontroller within commonplace cables used to implement the PUCK protocol and store the associated SensorML file in order to turn any commercial sensor into a PUCK-enabled plug-and-work unit. Figure 2 illustrates the high-level concept of an intelligent middleware layer.

The automated generation of EGO NetCDF files from vendor-specific data has been identified as another research challenge. Prototypical work on this topic has been realized within the Oceanids C2 project (Harris et al. 2019). It leverages metadata encoded as SensorML to establish an automated workflow for the creation of NetCDF files. Combining this approach with the ingestion into the presented Sensor Web architecture will lead to a sustainable solution for glider data management.

Acknowledgments

The work presented in this article was performed within the project BRIDGES as part of the European Union's Horizon 2020 research and innovation programme under grant agreement No. 635359.

References

Alseamer (2019): SEAEXPLORER. Accessed online on 20th February 2019: https://www.alseamar-alcen.com/products/underwater-glider/seaexplorer

Botts, M., Percivall, G., Reed, C. and Davidson, J. (2006): OGC Sensor Web Enablement: Overview and High Level Architecture. In: International conference on GeoSensor Networks (pp. 175-190).

Botts M. and Robin, A. (2007): OGC Implementation Specification: Sensor Model Language (SensorML) 1.0.0. OGC 07-000.

Braeckel, A., Bigagli, L. and Echterhoff, J. (2016): OGC Publish/Subscribe Interface Standard 1.0 - Core. OGC 13-131r1.

Bröring, A., Echterhoff, J., Jirka, S., Simonis, I., Everding, T., Stasch, C., Liang, S. and Lemmens, R. (2011): New Generation Sensor Web Enablement. Sensors 2011, 11, 2652-2699.

Bröring, A., Stasch, C. and Echterhoff, J. (2012): OGC Implementation Standard: OGC Sensor Observation Service. OGC 12-006.

Cox, S. (2011): OGC Implementation Specification: Observations and Measurements (O&M) - XML Implementation 2.0. OGC 10-025r1.

EGO gliders data management team (2017): EGO gliders NetCDF format reference manual version 1.2. https://doi.org/10.13155/34980

Harris, C., Lorenzo, A., Jones, O., Buck, J., Kokkinaki, A., Loch, S., Gardner, T. and Phillips, B. (2019): Oceanids C2. An integrated command, control and data infrastructure for the over-the-horizon operation of marine autonomous systems. Frontiers in Marine Science (pending publication).

ISO (2011): ISO 19156:2011 - Geographic information -- Observations and measurements - International Standard. Geneva, Switzerland, International Organization for Standardization.

Kokkinaki, A., Darroch, L., Buck, J. and Jirka, S. (2016): Semantically Enhancing SensorML with Controlled Vocabularies in the Marine Domain. In: Proceedings of the Geospatial Sensor Webs Conference 2016. Münster. 2016.

Kongsberg (2019): Seaglider AUV. Accessed online on 20th February 2019: https://www.km.kongsberg.com/ks/web/nokbg0240.nsf/AllWeb/EC2FF8B58CA491A4C1257B870048C78C?OpenDocument

Liang, S., Huang, C.-Y. and Khalafbeigi, T. (2016): OGC SensorThings API Part 1: Sensing. OGC OGC 15-078r6.

Schaap, D. and Fichaut, M. (2017): SeaDataCloud - Further developing the pan-European SeaDataNet infrastructure for marine and ocean data management. In: EGU General Assembly Conference Abstracts.

Teledyne (2019): Slocum G3 Glider. Accessed online on 20th February 2019: http://www.teledynemarine.com/slocum-glider

Toma, D. M., del Rio, J., Jirka, S., Delory, E., Pearlman, J. and Waldmann, C. (2015): NeXOS smart electronic interface for sensor interoperability. In: OCEANS 2015-Genova (pp. 1-5). IEEE.