

Collaborative Ocean Resource Interoperability: Multi-use of Ocean Data on the Semantic Web

Feng (Barry) Tao, Jon Campbell, Maureen Pagnani, and Gwyn Griffiths

National Oceanography Centre Southampton, UK
{bt, joc, gxg, mred}@noc.soton.ac.uk

Abstract. Earth Observations (EO) collect various characteristics of the objective environment using sensors which often have different measuring, spatial and temporal coverage. Making individual observational data interoperable becomes equally important when viewed in the context of its expensive and time-consuming EO operations. Interoperability will improve reusability of existing observations in both the broader context, and with other observations. As a demonstration of the potential offered by semantic web technology, we have used the National Oceanography Centre Southampton's Ferrybox project (where suites of environmental sensors installed on commercial ships collect near real time data) to set up an ontology based reference model of a Collaborative Ocean, where relevant oceanographic resources, such as sensors and observations, can be semantically annotated by their stakeholders to produce RDF format metadata to facilitate data/resource interoperability in a distributed environment. We have also demonstrated an infrastructure where common semantic management activities are supported, including ontology management, semantic annotation, storage, and reuse (navigating, inference and query). Once the method and infrastructure are adopted by other related oceanographic projects to describe their resources and move their metadata onto the semantic web, it would be possible to see better interoperability within the Collaborative Ocean initiative to facilitate multiuse of ocean data, as well as making more EO data available on the semantic web.

Keywords: Semantic Web, data/resource interoperability, oceanographic earth observation, Ferrybox, Multiuse of ocean data.

1 Introduction

Advances in Web and sensor technologies have made available more and more environmental observations, both remote and in-situ, accessible in real time and as archived datasets. For example, the Ferrybox project [1] at the National Oceanography Centre Southampton (NOCS) has been in operation for six years, making oceanographic observations from various "Ships of Opportunity". One of the Ferryboxes is installed in the engine room of the *Pride of Bilbao* passenger ferry that sails between the south of England and Northern Spain. The Ferrybox maintains a closed system where near-to-surface sea water is pumped through a set of sensors to carry out observations of various environmental parameters such as temperature, salinity, fluorescence (as an indicator of the quantity of chlorophyll), turbidity, dissolved oxygen and carbon dioxide. Since 2002 a subset of the data collected from the *Pride of Bilbao* has been automatically telemetered to computers at NOCS every 10 minutes. These data are used to generate near real-time, public web display of ship track and oceanographic measurement information [7].

Datasets are available for downloading and quality-controlled data are eventually archived at British Oceanographic Data Centre (BODC) to allow future data analysis requirement by environmental scientists to understand the change of our natural environment. The Ferrybox project observations are by nature, dense in time coverage but restricted to only a small spatial coverage for a set of limited measurements; Coarse but more global-scale measurements can be obtained through remote sensing observations, from regional aircrafts to earth-orbiting satellites. However, the lack of standardizations between these observations has made it difficult to collate measurements from multiple providers in order to automate intelligent processes, such as measurement cross-validation, data/services mash-up, and personalised knowledge representation based on intelligent processing of distributed measurement data. In other words, the lack of standards and semantics will "isolate important data streams and intensify the existing problem of too much data and not enough knowledge"[2].

On the other hand, as a multi-disciplinary subject, ocean science often involves collaborative work from different stakeholders and roles such as sensor development researchers, oceanographic instrument deployment engineers, data management, and environmental scientists, each of whom provides, as well as consumes data/services that are shared within the domain. Within the context of Oceans 2025, which is a strategic marine science programme funded by the UK Natural Environment Research Council (NERC), bringing together marine researchers to increase people's knowledge of the marine environment so that they are better able to protect it for future generations [13]. For example, in the Ferrybox project, engineers need to collaborate with the sensor providers in order to select and integrate sensors into the Ferrybox Instruments, which are then deployed on platforms (e.g. Ships of Opportunity in this case). Observations are sampled and brought back regularly by telecommunication through satellite and on-site visits to collect the physical memory storage. Real time monitoring is used to monitor the Ferrybox status. Calibrations for sensors are inevitably needed as the measurements drift due to bio-fouling and other reasons, hence their calibration history should be recorded to allow easy management for future reuse. Datasets of the observations are then archived by data managers with various metadata and descriptions attached so that domain scientists can investigate a particular environmental phenomenon, e.g. a bloom of plankton and its relationship with temperature and salinity [8]. All these activities through out the workflow have access to different yet related resources which can be annotated against a domain reference model to reduce ambiguity (every resource has a unique formal identifier) and improve connectivity (resources are linked to other resources through defined relations). This eventually will support sharing earth observation data and knowledge.

1.1 Related Work and Motivation

The Semantic Sensor Web (SSW) is a framework for providing well defined meaning using the semantic web technology in W3C that realises the idea of "having data on the Web defined and linked in a way that it can be used by machines not just for display purposes, but for automation, integration, and reuse of data across various applications" [4]. The underpinning modelling language for the semantic web is the RDF (Resource Description Framework), which is a directed-graph model using XML syntax that can also be expressed in many other forms, e.g. Notation3 (N3). The

RDF data model describes the real world by making statements in the form of *triples* (subject-predicate-object) where every item is either a literal or a URI (Uniform Resource Identifier) prefixed with a name space. RDF is an open system in that the namespaces can be unlimited, each one corresponding to an ontology that defines a formal conceptual representation within a domain - concepts and relationships in particular. The simple RDF data model allows computers to store, exchange, and to some extent reason (think!) using information from across the Web, providing opportunities for computer-assisted information handling with more assurance and certainty.

Semantic Web based knowledge management has been previously introduced and successfully applied in engineering design search optimization [14] and e-learning [15]. Both domains are single core inter-disciplinary, i.e. between computer science and one more area. The same application in environmental domain is multi-core inter-disciplinary area where the domain is more diversified to include for example oceanography, electronics, chemistry, biology, etc. Therefore setting up the infrastructure and gluing/reusing existing domain ontology are more important.

With these objectives in mind, we have created a Collaborative Ocean Group (COG) in the context of Marine Metadata Interoperability (MMI) [5] to demonstrate this proof of concept for multi-use of ocean data. In order to coordinate oceanographic activities and facilitate resource sharing and data interoperability, we describe in this article a semantic web infrastructure that helps different roles to semantically annotate their resources so that they can be managed as linked-data which are W3C compatible to facilitate semantic web processing.

2 Ontology Based Metadata Modeling

Metadata has been widely used in environmental data management systems to improve both the accessibility and the quality of the observational information. However, interoperability between these systems is often restricted due to the lack of semantics and closeness of their metadata reference models. In our work, we aim to improve the situation by adopting the Semantic Web approach. For example, the concepts of namespace and URI in the Semantic Web make it possible to describe resources in an open and globally accessible manner. Further more, recent developments in the Semantic Web community have made available more tools and APIs that can handle data on the semantic web for parsing, querying and inference.

2.1 An Ontology Based Reference Model for Collaborative Ocean

Ontology is a formal representation of a set of commonly agreed concepts and defined relationships between them. It provides shared but controlled vocabularies, makes domain assumptions explicit and can be used to describe resources in a domain in a way that all domain users can collaborate with each other (Figure 1 and Figure 2). In short, it is an accepted consensus of domain conceptualisation based on which semantics can be added to enrich domain resources.

RDF and OWL has been the main model for expressing ontology. The RDF data model is based on the idea of making statements about web resources in the form of subject-predicate-object triples, where each entity in the triple is either a literal or a RDF resource with a URI (Uniform Resource Identifier). A set of triples using RDF resources can be viewed as a directed RDF graph. OWL (Web ontology language)

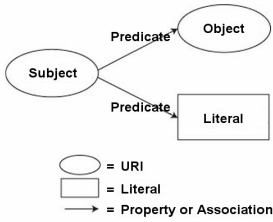


Fig. 1. RDF direct graph model

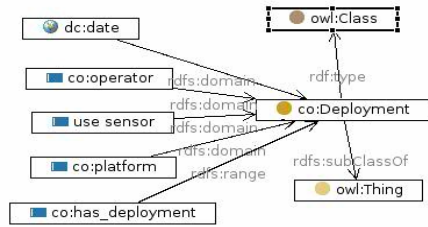


Fig. 2. An example in the Ferrybox project

extends the RDF Schema with additional vocabulary, e.g. on property restrictions (*allValuesFrom*) and restricted cardinality (*minCardinality*), etc.

The ontologies used can be categorized into three types according to their levels of generic description to the domain

Upper level and meta ontology

There are many generic conceptualisations that are domain independent and therefore can be used to provide cross-domain resource description.

DC - Dublin Core is a cross-domain XML/RDF based metadata standard designed to describe general media on the web to make them easier to find. It is divided into two levels - Simple Dublin Core (Title, Creator, Date, Subject, etc) and Qualified Dublin Core (Provenance and Rightsholders, etc). We use the DC to describe generic attributes of Ferrybox resources such as datasets collected and put online.

W3C RDF - defined internal vocabulary for RDF, such as *rdf:resource* and *rdf:datatype*, etc.

W3C RDFS - this is the root vocabulary and syntax specification originally defined in the RDF model. We use *rdfs:seeAlso* to maintain linkages between our parameters and BODC controlled vocabularies.

SKOS - noted as Simple Knowledge Organization Schema, it is designed to express generic relationship before concepts, such as *skos:broader*, *skos:related*, etc. BODC uses SKOS to express relationship among their control vocabularies.

Middle Level ontology

These ontologies are less generic than the upper level ontology but more generic than the overall domain ontology.

A common feature of EO data is its geospatial properties. Almost all EO measurements are related to location. We adopted the widely use W3C GEO vocabulary for this modelling requirement.

Another feature is the inclusion of social network related properties which are often required to describe the different roles in oceanography in such a way as to allow them to communicate and collaborate with each other in order to carry out EO tasks. We adopted the FOAF ontology that is originally proposed in W3C to describe social network related properties.

Domain ontology

Domain ontology is specific to the domain application and is normally non-transferring for other different domains. We build the CO ontology by investigating the domain.

CO - Collaborative Ocean ontology

This is the main specific reference model ontology designed to provide domain conceptualisation needed for describing domain related resource annotations. The ontology has the namespace “co” and focuses on concept taxonomy and how they are connected to other concepts through defined properties (Figure 3).

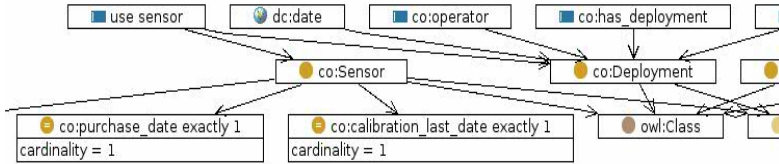


Fig. 3. A section of the CO ontology

Some of the important concepts are Sensor, Platform, Deployment, Datasets, etc. RDFS and OWL has been used to explicitly describe concept taxonomy, relationship and constraints in term of other define properties.

2.2 Interoperability with Existing Reference Model

Interoperability with other reference models such as MMI device ontology and BODC vocabulary is maintained by crossed linkage and references. One example is shown below where we have reused BODC measurement parameter vocabulary by adding SKOS annotations. This would provide the semantic linkage between two reference models when navigating on the semantic web from one to the other.

In parallel with the development of the Collaborative Ocean ontology, we have also participated in the MMI Device Ontology Working Group, with an aim to contribute our experience and requirement for sensor annotation metadata and to eventually reuse

```

<co:ChemicalParameter
rdf:about="http://www.soton.ac.uk/~bt/ontology/ocean/collaborativeoce
an_ins.rdf#salinity">
  <rdfs:seeAlso
rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
>http://vocab.ndg.nerc.ac.uk/term/P021/25/PSAL</rdfs:seeAlso>
  <co:unit
rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
>u2030</co:unit>
<co:unit
xml:lang="en">ppt</co:unit>
</co:ChemicalParameter>
    
```

* RDF snippet that links to BODC vocabulary

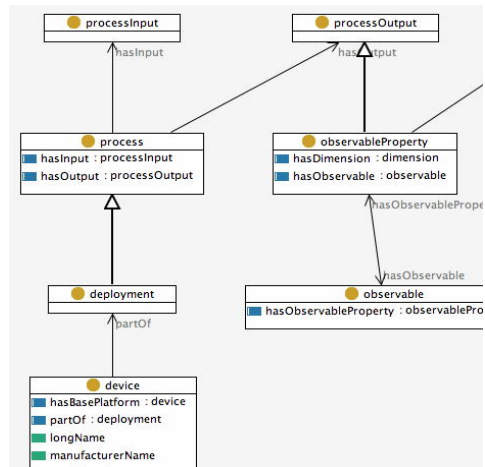


Fig. 4. MMI Device ontology

their device ontology in the CO ontology. This will provide end users with more choices to carry out their sensor device semantic annotation and better enable interoperability in a broader EO context. Figure 4 is a snapshot of the MMI device ontology concept diagram in development.

3 A Semantic Web Infrastructure for the Collaborative Ocean

In order to efficiently describe, store and reuse the oceanographic EO operational resource and measurement data on the semantic web, we have set up a layered infrastructure as shown in Figure 5. An advantage of this structure is to allow distributed management of the semantic web information at different granularities, and to expose a different level of technical details to different roles through customized interfaces and toolkits.

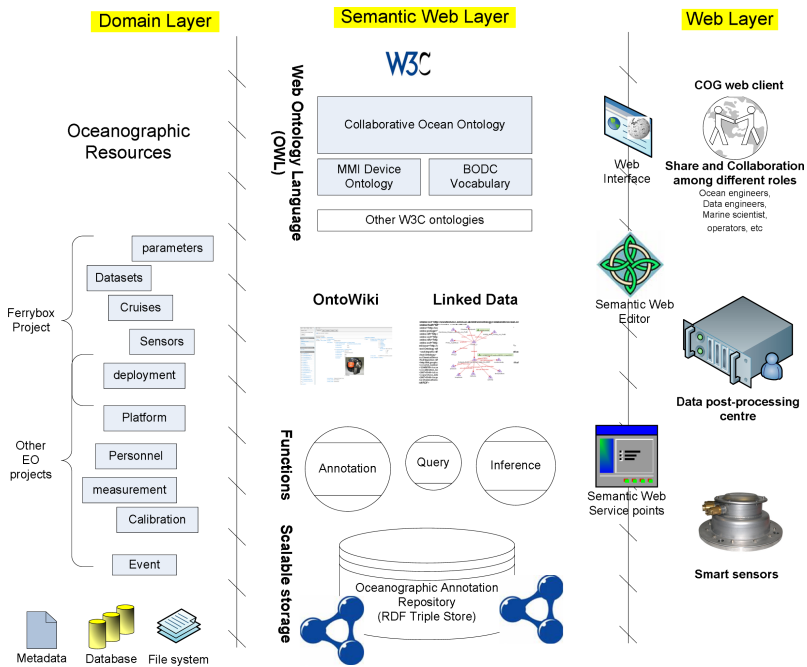


Fig. 5. Layered diagram of the semantic web infrastructure

Domain layer

Domain resources such as measurement datasets, calibration reports, and sensor profiles as well as existing metadata are stored in this layer, in the form of flat files or database records. The Collaborative Ocean Ontology and relevant controlled vocabulary are modelled and adopted based on study of the resources in this layer as well as best practice directly from experts operating the Ferrybox project. We use not

only existing metadata such as Dublin Core and FOAF but also related domain vocabulary such as MMI device ontology and BODC vocabulary. The CO ontology serves as the domain reference model that provides important concepts and assumptions of the domain resource conceptualisation to allow proper semantic annotation of the Ferrybox resources and their publishing on the semantic web. Collaborative Ocean ontology has been built using Collaborative Protégé [9], which is an extension of the existing Protégé system that supports collaborative ontology editing. The Ontology is further maintained in the Topbraid Editing environment [10] to exploit its advantage in managing multiple ontology namespaces and its RDF triple representation of the ontology as well as the semantic annotations simulation. The advantage of Topbraid composer over protégé is its capability in managing multiple external namespaces and accessibility to the openRDF triple store.

During the Knowledge Acquisition (KA) process, we have adopted an agile knowledge modelling method that encourages domain experts to participate in the KA as an iterative process. A Wiki page has been set up to assist recording sensor profile text as well as key description of resources available in the Ferrybox project. A Protégé server has been setup to drive a collaborative protégé client to collect concepts and resource taxonomy from various domain experts in the group for a very initial version of the CO ontology. The RDF data is then fed into an OntoWiki [11] web based knowledge acquisition system as an information map where intuitive authoring of semantic RDF content as well as simulation of the semantic annotation process is made possible to enhance understanding of the domain. The OntoWiki adopted an agile and adaptive knowledge engineering strategy that is similar to a wiki-style paradigm of “making it easy to correct mistakes, rather than making it hard to make them” [12]. The Semantic Wiki is another similar tool that integrates the Semantic Web into a Wiki system. Interested users can find a recent review at [18].

Semantic Web layer

This is the layer where semantic web based functionalities such as domain resource annotation, browsing, query and inference are made available as services though integrating existing semantic web APIs with the ontologies and domain resources in the domain layer. The layer is driven by the CO ontology from the domain layer. We use Jena Semantic Web API for local semantic processing and remote access to a Sesame triple store, which is a persistent storage of all the semantic annotations in RDF triple format. The HP Jena Semantic Web API is so far the most comprehensive java API to manipulate RDF models while the sesame openRDF delivers a better semantic web framework. Several methods have been used to generate and manage these RDF triples, namely the Collaborative Protégé, OntoWiki, TopBraid and a COG web client.

The Topbraid composer is a flexible and versatile semantic web editing tool that can manage the semantic web triples in different formats such as illustrated in Figure 6, where relevant RDF triples can be collected through standard SPARQL query to the Sesame triple store and presented as a directed linked graph as well as XML based RDF files.

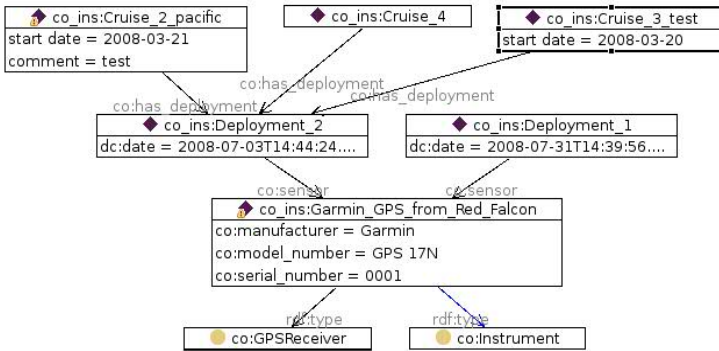


Fig. 6. RDF graph representation of the triples

Similar semantic web annotation and query functionalities are also interfaced to a COG web client (at the Web layer) using Java Server Face technology to allow users with less semantic web experience to contribute and reuse semantic web metadata on oceanographic observational resources.

Web layer

This is where applications are deployed to interface end users with the functionalities offered by the semantic web layer in the context of the domain. We have implemented a COG web client that allows manual and semi-automated semantic management of oceanographic resource share and collaboration using semantic web functionalities offered in the semantic web layer of the system. More intelligent, semantic web driven operations, such as data post-processing and smart sensor activities can be also supported in the future.

4 Technical Implementation and How the System Works

The system components/services have been deployed onto multiple nodes of different operating systems (Windows XP and Linux) to demonstrate a distributed Service Oriented Architecture (SOA). As shown in Figure 7, we have chosen to use Java and PHP as the main development languages due to the fact that most semantic web APIs (such as Jena and Sesame) are written in Java and its portability in different systems. PHP is another popular Web language especially suitable for large scale distributed systems across the Web. We have used Apache and Tomcat to run the web server and provide a Java servlet container where various services are deployed, such as an ontology driven Java Facelet web interface service that invokes semantic web functionalities such as query and inferences to interact with the semantic information in the RDF triple store. The triple store is a Sesame openRDF repository package deployed in another node across the network, where OntoWiki is also deployed to offer an alternative access to the semantic information.

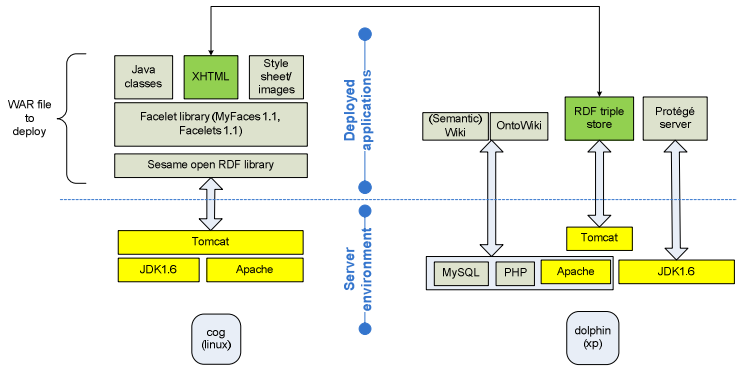


Fig. 7. Technical architecture of the system

Connecting to the system

A Sesame server address is required to specify the location where all the semantic annotations are stored as RDF triples. Multiple repositories can co-exist in the same server to store the RDF triples.

The user can also indicate whether inference is enabled. The inference is at the RDFS level to include inferred new triples according to RDF schematic reasoning rules. More interesting domain rules can be designed to produce inferred triples based on domain business logics.

Server: http://cog.noc.soton.ac.uk:8080/openrdf-sesame/repository/co_all

PLEASE CONFIGURE SERVER ADDRESS AND REPOSITORY ID BELOW

SesameServer:

RepositoryID: (try co_ontology or co_instance or co_all)

Inference?:

Fig. 8. Connecting the RDF triple store

Dynamic ontology driven semantic annotation of domain resources

Most initial semantic annotations are entered through other channels by our domain experts using Protégé and Topbraid Semantic web toolkits. Once they are stored as RDF triples in the repository, other interested users will be able to reuse part of these triples to generate further semantic annotations in a web interface. Meanwhile, reasoners can be used to generate further RDF statements that represent inference closure results of existing triples.

The web based semantic annotation page is dynamically generated by querying the CO ontology using SPARQL, therefore it is adaptive during the maturing the CO ontology. SPARQL is the querying language standard for collecting data from the semantic web. It is similar to the SQL for database, but with added semantic web

features based on the RDF model and multi-source RDF query across the Web. Below is a SPARQL query for defined properties of the domain class *co:ChemicalSensor*, used to dynamically generate the semantic annotation forms as shown in Figure 9.

```
SELECT
DISTINCT ?property
WHERE { {?property
rdfs:domain ?class_X }
{co:ChemicalSensor rdfs:
subClassOf ?class_X }
UNION {?property
rdfs:domain
co:ChemicalSensor} }
```

Fig. 9. Ontology driven semantic annotation web form generation

With the ontology driven annotation form, users can annotate datasets by providing values for a set of properties as defined in the ontology. These annotations will be converted as RDF triples and stored in the semantic web triple store so that other users can access through further query or browsing.

Semantic search, browsing and navigation

Semantic search, browsing and navigation of the annotations can also be achieved through submitting SPARQL semantic queries to the triple store and rendering results as linked data [19] on the web. Users will be able to view a list of semantically enriched instances of a particular class. They can click on any interesting element of the semantic annotations, for which the system will generate a new query accordingly, and get back another set of matched triples for rendering on the web. As all the resources can be annotated by their stakeholders at any time, these changes will be reflected in other semantic web management activities that reuse these triples.

An example of the semantic search and navigation is illustrated in Figure 10. According to the semantic annotations, the Ferrybox deployments are associated with the “*co:Cruise*” concept where additional properties such as “*co:aboutSea*”, “*co:startPort*” and “*co:dataset*” can be assigned with values during the semantic annotation process. The semantic query retrieves a list of annotated cruise instances on which users can further browse their semantics and navigate to other linked semantic annotations.

Similarly, queries to dataset annotations can help users to discover their linkages to other annotated resources in the reference model, such as measurement parameters, original dataset producers and visualization plots generated, etc. Figure 11.

A notable feature is the resolvable namespace prefixes that point to corresponding published ontologies that define the underlying semantics.

Server: http://coig.nrc.soton.ac.uk:8080/openrdf-sesame/repository/co_owl

FIND SEMANTICALLY ANNOTATED INSTANCES:

Available classes:

co:Cruise

co_ins:Cruise_1_pride_of_Bilbao
 co_ins:Cruise_2_pacific
 co_ins:Cruise_3_test
 co_ins:Cruise_4
 co_ins:Cruise_port_test

▪ **co_ins:Cruise_1_pride_of_Bilbao**

dc:description	"this is a text description test"^^xsd:string
smil:selectQuery	"SELECT ?s ?p ?o WHERE { ?s rdfs:subClassOf ?o }"^^xsd:string
co:aboutSea	co_ins:Sea_2
co:aboutSea	co_ins:Sea_2_Irish_Sea
co:dataset	co_ins:Dataset_Pride_of_Bilbao_data_for_2007
co:endPort	co_ins:bilbao_port
co:has_deployment	co_ins:Deployment_1
co:startDate	"2008-04-09"^^xsd:date
co:startPort	co_ins:Port_1_Portsmouth
sparql:query	"SELECT ?s ?p ?o WHERE { :Cruise_1_pride_of_Bilbao ?p ?o }"^^xsd:string
rdfs:type	co:Cruise
rdfs:comment	"test"^^xsd:string

Other related triples

```
co_ins:Deployment_1 co:sensor co_ins:Aanderaa_optode_3835_33
co_ins:Deployment_1 co:sensor co_ins:Garmin_GPS_from_Red_Falcon
co_ins:Deployment_1 co:sensor co_ins:Main_flow_sensor_on_PoFb_ferrybox
co_ins:Deployment_1 co:sensor co_ins:SeaBird_Hull_temperature_5BE480002
co_ins:Deployment_2 co:sensor co_ins:Aanderaa_optode_3835_641
co_ins:Deployment_2 co:sensor co_ins:Garmin_GPS_from_Red_Falcon
co_ins:Deployment_2 co:sensor co_ins:PressureSensor_in_Chelsea_Minipack
```

All namespaces

```
dc http://purl.org/dc/elements/1.1/
rdfs http://www.w3.org/2000/01/rdf-schema#
co http://www.soton.ac.uk/~bt/ontology/ocean/collaborativeocean.owl#
```

Fig. 10. Semantic browsing of Cruises

Mashing up with OGC service

Since the establishment of the Open Geospatial Consortium (OGC) in 1994, a lot of standards and services have been developed to coordinate the geospatial industry all over the world. One of the most relevant initiatives in the OGC is Sensor Web Enablement (SWE), which aims to enable real-time integration of heterogeneous sensor observations into the Web infrastructure. The main specifications include Sensor Model Language (SensorML), Sensor Observation Service (SOS), Sensor Planning Service (SPS), etc [6]. Various XML schemas are defined in SWE to accommodate XML files used in service in/output, e.g. in SOS, SensorML and Transducer Markup Language (TML) are used in *describeSensor* method response. The *getObservation* method has an input XML schema that uses Geographical Markup Language (GML) to specify time-spatial coverage of the observations. Observation offerings in SOS also provide non-overlapping observations constrained by a number of parameters such as geographical region and phenomena being sensed. This allows easy integration of spatially referenced data to layers in Web Map Service (WMS), another important OGC specification.

FIND SEMANTICALLY ANNOTATED INSTANCES:

Available classes:
co:Dataset

Submit Query

co_Ins:Dataset_Pride_of_Bilbao_data_for_2007
co_Ins:Dataset_test
co_Ins:pacific_celin_data_2008

co_Ins:Cruise_1_pride_of_Bilbao co:dataset
co_Ins:DataLogging_1_example co:datasetOutput

co_Ins:Dataset_Pride_of_Bilbao_data_for_2007

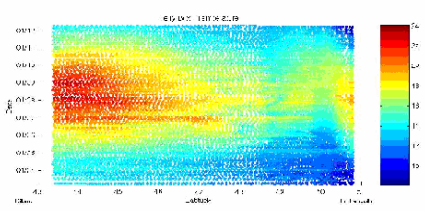
co:aboutParameter co_Ins:conductivity
co:aboutParameter co_Ins:latitude
co:aboutParameter co_Ins:longitude
co:aboutParameter co_Ins:pressure
co:aboutParameter co_Ins:salinity
co:aboutParameter co_Ins:temperature
co:DataLocation "S:\OBEPRIV\Ferrybox\BODC_Data\2007"^^xsd:string
co:has_visualization co_Ins:DataPlot_Fluorescence
co:has_visualization co_Ins:DataPlot_Salinity
co:has_visualization co_Ins:DataPlot_Temperature
rdf:type co:Dataset

Other related triples

co_Ins:Cruise_1_pride_of_Bilbao co:has_deployment co_Ins:Deployment_1
co_Ins:Cruise_2_pacific co:has_deployment co_Ins:Deployment_2
co_Ins:Cruise_3_test co:has_deployment co_Ins:Deployment_2
co_Ins:Cruise_3_test co:has_deployment co_Ins:Deployment_3
co_Ins:Cruise_4 co:has_deployment co_Ins:Deployment_2

All namespaces

dc http://purl.org/dc/elements/1.1/
protege http://protege.stanford.edu/plugins/owl/protege#
sm http://topbrail.org/sparqimobon#
geo http://www.w3.org/2003/01/geo/wgs84_pos#
co_Ins http://www.soton.ac.uk/~bt/ontology/occean/collaborativeocean_Ins.rdf#
foaf http://xmlns.com/foaf/0.1/



The heatmap shows a grid of data points with a color scale on the right ranging from 10 (blue) to 25 (red). The x-axis is labeled 'Longitude' and ranges from -14 to 10. The y-axis is labeled 'Latitude' and ranges from 0.000 to 0.010. The data shows a clear pattern of higher values (red/orange) in the lower-left quadrant and lower values (blue) in the upper-right quadrant.

Fig. 11. Semantic browsing of datasets

In the Ferrybox context, we have done some experiments to demonstrate that some resource annotations, such as those of the sea ports, which have been semantically annotated using geospatial metadata, can be queried and their results mashed up with third-party WMS services such as Google map to visualize their geospatial presence. This demonstrates interoperability potential between RDF triples and OGC standard. We believe that as more measurement data are annotated with geospatial references, the mashing up will provide more interesting integration between the W3C semantic web and the OGC services.

5 Discussion and Future Work

The Semantic Web technology has moved from a pure research topic to more and more applications in different domains [15] [16] thanks to the growing availability of tools and implementation of W3C specifications. The bottle-neck that prevents abundant publishing of semantic content still lies in the semantic annotation. This is because users tend to hesitate in carrying out the somewhat tedious semantic annotation tasks without being able to see immediate benefit of the input. We believe that this problem can be relieved through two approaches - distributing the task load and batch data conversion.

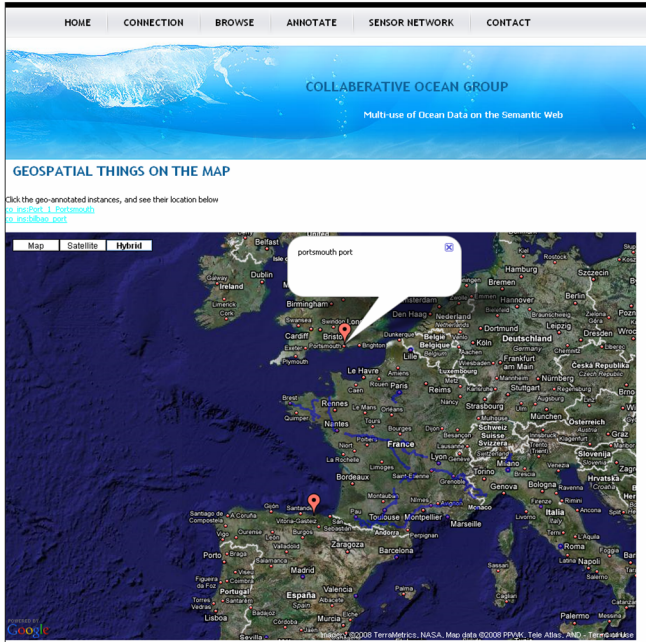


Fig. 12. Query and plotting of spatial annotations on the Google map

Since 2002 the data collected from the Pride of Bilbao Ferrybox every 10 minutes has been stored in a database which is used to generate annual measurement graphs shown on the web. However, users or software agents reading the web site have no access to formal descriptions of other related resources that actually collect the measurement datasets. It would be interesting to access such semantic information, e.g. sensor profiles, their calibration and deployment history, from the semantic annotation repository, as they are annotated by their stakeholders in the RDF triple store. This could also help the user to find other similar EO data for comparison by match-making and reasoning their semantic annotations on certain aspects of ontological reference model. We also plan to extend the system to a broader context in EO to incorporate more resource annotations into the system.

As can be seen in this paper, there is a clear distinction between ontology and semantic annotation, both of which can be represented in RDF. The ontology, which contains higher level schema-like domain knowledge, has more restricted written access once it reaches a finalized version and is published at a URL. It then serves as a set of formal reference points for definition of all necessary semantics used in the semantic annotations. The semantic annotations are however stored in the triple store which allows more flexible management including modifying, appending, and removing of the RDF triples. This is called the Semantic Web approach. One the other side, for a more open system with less control on the domain schema, ontology can be viewed as having the same level as semantic annotations. That means both of them are stored in the RDF triple store without real distinctions. In this way the ontology

can also be modified by end users just as they can do with the semantic annotation triples. This is called the semantic web (lower case) approach. We use the first approach in our work but plan to try the second approach as well for comparison.

With the initial COG infrastructure set up at NOCS, it is now possible to semantically enrich oceanographic resources with ontology based metadata, either defined locally or externally. We have encouraged engineers at NOCS to contribute semantic annotations of their resources in a distributed way and the resulting RDF triples are stored in a persistent storage so that they can be reused by other engineers and data management applications in a later phase.

6 Conclusion

In this paper we presented our work on Collaborative Ocean Initiative using the Ferrybox project as a context. The aim is to facilitate resource and data interoperability in oceanographic Earth Observation (EO) to allow multi use of the ocean data. We investigated various technologies and methods, in particular the semantic web approach. By following a semantic web based knowledge life cycle, we have carried out activities on ontology modelling, semantic web annotation of domain resources and their storage as RDF triples and publishing/reuse on the. A semantic web infrastructure has been set up to accommodate these activities. Based on domain knowledge acquired, an ontology based domain reference model has been developed to provide domain conceptualisation and assumptions in the Ferrybox project. Through using different Semantic Web tools and interfaces, we have semantically annotated operational resources such as sensor and platform, as well as observation datasets in the Ferrybox project on the ontology based domain reference model. The ontology and semantic annotations are stored as OWL and RDF triples in a semantic web triple store where they can be further reused to facilitate dynamic ontology driven semantic annotation, linked data browsing, query and data integration.

We have learned through these semantic web activities that the data interoperability problem in EO can be improved by adding a semantic web layer where the ontology based domain reference model is used to enrich EO resources with standard metadata shared within the community.

Acknowledgements

This study is supported by the UK Natural Environment Research Council (NERC) through the Oceans 2025 core programmes of the National Oceanography Centre, Southampton.

References

1. Hydes, D.J., Campbell, J.M., Dunning, J.: Systematic Oceanographic Data Collected by FerryBox, Sea Technology (February 2004)
2. Semantic Sensor Web at Wikipedia,
http://en.wikipedia.org/wiki/Semantic_Sensor_Web2

3. Sheth, A., Henson, C.: *Semantics & Services - Semantic Sensor Web*. IEEE Computer Society, Los Alamitos (2008); 1089-7801/08, IEEE INTERNET COMPUTING, Kno.e.sis Center, Wright State University
4. The World Wide Web Consortium (W3C), <http://www.w3.org>
5. MMI-COG, <http://marinemetadata.org/community/teams/cog>
6. OpenGIS® Sensor Observation Service Implementation Specification, OGC 06-009r1, 2006-02-13
7. Ferrybox real time monitoring, http://www.noc.soton.ac.uk/ops/ferrybox_index.php
8. Hydes, D.J., Hartman, M.C., Barger, C.P., Campbell, J.M., Cure, M.S., Woolf, D.K.: A study of gas exchange during the transition from deep winter mixing to spring bloom in the Bay of Biscay measured by continuous observation from a ship of opportunity. *Journal of Operational Oceanography* 1(2), 41–59 (2008)
9. Tudorache, T., Noy, N.F., Tu, S.W., Musen, M.A.: Supporting collaborative ontology development in Protege. In: Sheth, A., et al. (eds.) ISWC 2008. LNCS, vol. 5318, pp. 17–32. Springer, Heidelberg (2008)
10. TopBraid Composer, <http://www.topquadrant.com/topbraid/composer/index.html>
11. Auer, S., Dietzold, S., Riechert, T.: OntoWiki – A Tool for Social, Semantic Collaboration. In: Cruz, I., Decker, S., Allemang, D., Preist, C., Schwabe, D., Mika, P., Uschold, M., Aroyo, L.M. (eds.) ISWC 2006. LNCS, vol. 4273, pp. 736–749. Springer, Heidelberg (2006)
12. Leuf, B., Cunningham, W.: *The Wiki Way: Collaboration and Sharing on the Internet*. Addison-Wesley Professional, Reading (2001)
13. Ocean2025, <http://www.oceans2025.org/>
14. Tao, F., Shadbolt, N., Chen, L., Xu, F., Cox, S.: Semantic Web based Content Enrichment and Knowledge Reuse in e-Science. In: Meersman, R., Tari, Z. (eds.) OTM 2004. LNCS, vol. 3290, pp. 654–669. Springer, Heidelberg (2004)
15. Tao, F., Millard, D., Zalfan, M., Chen, L., Davis, H.: Knowledge based Learning Experience Management on the Semantic Web. In: IADIS International Conference of e-Learning, Lisbon, Portugal, July 6-8 (2007)
16. Lewis, M., Cameron, D., Xie, S., Arpinar, B.: ES3N: A Semantic Approach to Data Management in Sensor Networks. In: *Semantic Sensor network workshop, the 5th International Semantic Web Conference, ISWC 2006, November 5-9, Athens, Georgia, USA* (2006)
17. Berners-Lee, T., Hendler, J., Lassila, O.: The Semantic Web. *Scientific American*, 34–43 (May 2001)
18. Fowler, D., Guan, T., Tao, F., Wills, G., Crowder, R., Shadbolt, N.: Comparison of cfd wikis. Technical Report at School of ECS University of Southampton (2008)
19. W3C Linked Data, <http://www.w3.org/DesignIssues/LinkedData.html>