

Semantic Solutions for Integration of Federated Ocean Observations

M. A. Cameron, Jemma X. Wu, Kerry Taylor, David Ratcliffe
Geoffrey Squire, and John Colton

CSIRO ICT Centre, Australia

Abstract. Ocean observing is critical to the design and development of ecosystem-based management of public health risk, water quality, and living marine resources. Responsibilities and resources for observing the ocean, as for other global natural resources, are distributed across regional, national and international agencies, with decentralised authority and limited funding. We discuss the challenges for shared access to sensor-derived observations in the context of the Integrated Ocean Observing System (IOOS), an initiative of the US National Oceanic and Atmospheric Administration (NOAA). The system aims to federate heterogeneous regional ocean observations. We highlight three technical approaches to integration; the first based on the Open Geospatial Consortium's (OGC) Sensor Web Enablement (SWE) technology. The second, the OpenIOOS Portal, builds on that foundation but adds a measure of semantic interoperability. The third approach, our own Semantic Service Architecture (SSA), is also built upon SWE technologies but makes rich use of complex ontologies and mappings to offer a more flexible, semantically-driven integration platform. We analyse the three approaches for their ability to serve the needs of the ocean-observing community. We conclude that the latter rich-semantics approach is the most effective but requires a more advanced semantic literacy of the ocean observation community to gain the full benefit.

Key words: federated information systems; Sensor Web Enablement; Ocean Observing; Semantic interoperability

1 Introduction

Ocean observing is critical to the design and development of ecosystem-based management of public health risks, water quality and living marine resources. Thousands of observing systems are operating in the water. These systems include real time and delayed sensors that collect a wide variety of ocean observations, biological, chemical, geological and ecological properties. The user community involves scientific research agencies, industry groups, private enterprises, government regulators and policy-makers. Observations are made by a wide variety of sensors which are owned and operated by organisations including federal government, regional and state governments, non-government agencies,

private sector enterprises, and academic institutes. These organisations manage and monitor the sensors they deploy and have traditionally developed individual approaches to management of the data arising from the sensor observations to suit their own goals. This variety introduces obstacles to data access across the community, which are detrimental to the common goals of the organisations. The diversity and heterogeneity of the ocean observing systems obstructs the information flow needed to support decision making and promotion of economic, environmental and social benefits.

The U.S. National Oceanic and Atmospheric Administration (NOAA) has proposed a vision for a U.S. Integrated Ocean Observing System (IOOS)¹, to seamlessly integrate many of these ocean observing systems. IOOS represents a partnership of 17 Federal agencies and 11 Regional Associations (RA) sharing responsibility for the design, operation, and improvement of the national network of observations. Participants include NOAA's National Data Buoy Center (NDBC) and the Northwest Association of Networked Ocean Observing Systems (NANOOS) RA. The goal of IOOS is to continuously provide timely quality controlled data and information on the past, current and future states of the oceans from the global scale of ocean basins to the local scale of coastal ecosystems.

The IOOS has three components or subsystems: Observing subsystem, Data Management and Communications (DMAC) subsystem and the Modeling and analysis subsystem. Each subsystem faces different challenges. In this paper, we focus on the DMAC subsystem, although we also suggest a route towards integration of the modeling and analysis facilities with the DMAC subsystem.

One way to manage the diversity of the ocean observing systems is to develop and implement standards for the representation of data and its associated meta-data. IOOS has been considering adopting many current open standards such as the Open Geospatial Consortium's (OGC) Sensor Web Enablement (SWE) package. The SWE standards define both an interaction protocol and an XML data representation format suitable for a Service Oriented Architecture-styled system for discovery and retrieval of observation data. Experimental implementations of the SWE Sensor Observation Service have been deployed and are discussed in detail in this paper.

However, the SWE standards only provide a syntactic format for data representation. The shortcomings of this have been observed by the OGC community [1] and the broader research community [2]; proposed enhancements such as the OpenIOOS portal [3] have been developed. The portal provides a Web interface that permits search by observed property over a large number of SOS services. A semantic mapping is introduced that enables several similar observed properties to be searched via a single encompassing concept.

In our own work, we have experimented with a semantic web service composition approach using a Semantic Service Architecture (SSA) platform [4]. The SSA platform is semantic-intensive: relying on rich service descriptions expressed as an ontology that is related to underlying services through expressive mappings [5]. We have configured the SSA for the IOOS problem, in a tool we

¹ <http://ioos.gov/>

call the SIOOS. The SIOOS enables a user's expressions over a domain ontology to be resolved to service compositions that may include both data-oriented services like SOS and other modelling and analysis tools into a single executable workflow. Compositions developed using SIOOS may be deployed as fresh services for the community, complemented with suitable client interfaces for casual users.

In this paper, we analyze these three approaches to the ocean observation interoperability problem. They all rely on data presented through the underlying SOS services, but they differ chiefly in their levels of exploitation of semantic representations. The first approach uses SOS directly with no semantic tools. The second approach uses SOS enhanced with a low-level of semantic integration, specifically a terminology mapping, for discovery purposes. The third approach implemented by our SIOOS prototype, exploits rich semantic definitions of the underlying data to enable composition of services and integration of data. We claim that semantics can greatly improve the capabilities of ocean observation interoperability through these technologies.

In the next section of the paper we discuss the technical challenges faced by the IOOS federation in more detail. We follow this by a presentation of each approach in order of increasing semantic exploitation in sections 3, 4 and 5, and then discuss how well each approach meets the challenges in section 6 before concluding.

2 Challenges for IOOS

A distributed federation of independently owned and operated network-accessible services and datasets presents many opportunities and challenges to both information consumers and service providers. A regionally-based network of participants allows the federation's service providers to maintain focus on servicing local requirements and goals, while contributing to broader federation goals with a small additional overhead. However, this advantage at the regional level may conflict with the federated goal because uniformity of data collection, representation and distribution is emphasised at the regional level and de-emphasised at the federated level. Tradeoffs in quality of service at each level are inevitable.

In this section we discuss the technical challenges faced in order to achieve the benefits expected of a successful implementation of the IOOS, covering the needs of both the users of the system and the data contributors. Our challenges will provide a checklist against which our three solutions may be assessed later in the paper.

Facilitates data access. A first goal of the federation is to facilitate access by users to sensor observations. How well does the technology serve to publish data and to enable its discovery by potential users of the data?

Facilitates data discovery and understanding through contextual placement. The final goal of sensor deployment is to enable users to better understand the environment. A user seeks sensor data because they believe that some observa-

tions may help to solve a problem in his or her disciplinary domain. In a multi-disciplinary user community, users need data that is contextualised within their own problem domain in order to discover it using the terms of their domain and to assist in interpreting what data they can find. Does the solution assist problem-specific data discovery and interpretation?

Facilitates sound application of data through integration and interpretation. Beyond enabling data access, support for understanding, interpretation and application of the data is desirable. Usually, raw observation data needs processing by analytical tools to reach a human-interpretable form. For example, measurement of sea surface temperature at fixed points may be better understood as an interpolated contour map. Also, the interpretation of one phenomenon may require correlation from various sensor observations. For example, estimations of lobster catch require correlation of information about ocean depth, currents, and temperature [6]. Does the solution streamline the method for processing and integrating information from heterogeneous sensor platforms for a user?

Supports autonomy (of service provision by providers and regional solutions). The independent nature, divergent missions and investment priorities of organizations participating in the federation make it difficult to establish and maintain uniformity: at the level of standards implementation because of heterogeneous computing infrastructures; at the level of robustness and operational reliability of services because of diversity in organizational mission (for example NANOOS and NDBC); at the level of security; data quality; metadata; and configuration management. Does the solution allow providers to contribute at various levels of commitment while supporting users to make use of what is provided? To what extent does the solution allow regions to do things in their own way, while servicing the federation at little extra cost?

Low cost (of acquisition and maintenance). Organisations such as members of the National Federation of Regional Associations for Coastal and Ocean Observing² generally can not afford to invest heavily in data-sharing initiatives: cooperation with other agencies is generally a low priority goal despite the well-recognised mutual benefit; and IT budgets are always tight because IT is far from core business of the organisation. Long-term low cost of maintenance is even more important to such initiatives, because, although acquisition funding maybe granted through special project funds; maintenance and enhancement usually has to be funded through the agencies' own recurrent funding program. Can the solution be implemented and maintained at low cost?

3 SWE technology: integrating ocean observations with no semantics

This section introduces and discusses the first approach to integrating ocean observations using Sensor Web Enablement (SWE) technology [7]. This is a

² <http://www.usnfra.org/>

non-semantic based approach and forms a foundation for the remaining semantic approaches.

SWE is a working group of the OGC. The goal of SWE is to specify interoperability interfaces and encode metadata to enable integration of heterogeneous web-connected sensors. SWE has developed a number of standards including Observations & Measurement (O&M), Sensor Model Language (SensorML), Sensor Observation Service (SOS) and so on. Users can access observations from heterogeneous sensors by using platforms and products that implement these standards. This standard improves interoperability and integration of the heterogeneous sensor resources. Some organizations in the IOOS federation have deployed SWE servers for ocean observing sensors.

SOS defines document-based APIs for discovery and retrieval of sensor observations. SOS operations are classified into three profiles: core, transactional, and enhanced. The core profile consists of three mandatory operations (GetCapabilities, DescribeSensor and GetObservation). The other two profiles are optional. In this paper, we limit our scope to the core profile, as most SOS implementations do. With SOS services, users can query and retrieve sensor observations through a web client, following the publish-find-bind service-oriented architecture.

The GetCapability operation allows a user to retrieve service metadata about a specific SOS service instance. The response to a GetCapability operation is an XML document which is called the capability document. The capability document contains service metadata entries such as ServiceIdentification, ServiceProvider, ObservedProperty, FeatureOfInterest, and ResultFormat. These entries, together with optional or mandatory flags for entry values, provide clients with the grammar and parameter values required to construct DescribeSensor and GetObservation requests. The SOS standard requires a mandatory HTTP Post service end-point in which case the request is an XML document; however the IOOS federation also supports service endpoints where the request is a CGI query string.

While SWE does not specify a client interface, a simple UI client allows a user to undertake the steps to obtain observation data from a SOS server. The client first sends a GetCapability request to a SOS server and get its capability document. The client then analyses the capability document to find various parameters of an SOS offering. These parameters include the offering identification, observed property and response format. To invoke the getObservation operation, the client needs to form a query which looks like the query in Example 1.

Example 1. Here is a SOS getObservation query which specifies to get the “WaterTemperature” observations from the NDBC SOS with the response format schema to be “ioos/0.6.1”.

```
http://sdf.ndbc.noaa.gov/sos/server.php?VERSION=1.0.0&SERVICE=SOS&
REQUEST=GetObservation&offering=urn:x-noaa:def:station:noaa.nws.
ndbc::21413&observedProperty=http://www.csc.noaa.gov/ioos/schema/
IOOS-DIF/IOOS/0.6.1/dictionaries/phenomenaDictionary.xml%
23WaterTemperature&responseFormat=text/xml;schema="ioos/0.6.1"
```

The client will receive an XML document containing the data in response. A user who wants to get the observation data for water temperature from all the SOS installations in the IOOS, will have to repeat this interactive procedure for each server, and manage each of the response data documents.

Standardizing the sensor observation access interface does not really solve the interoperability problem between different IOOS entities. Even though most of the IOOS entities have adopted the OGC SOS standard when developing and deploying their own SOS web services, there remains lots of diversity in the various IOOS SOS services. One reason is that, like many other standards, SOS has been evolving since it was proposed in 2002 and several versions of SOS have been published. From our experience in using the IOOS SOS services, we found that there is wide variation in the different versions of SOS schemas. The other reason is that the SOS only standardised the schemas or the parameters but not the parameter values. Thus, different SOS implementations may use different terminology for the same observed property or phenomenon. For example, some IOOS SOS implementations use “WaterTemperature” while others use “sea_water_temperature”, or “WATER_TEMP” to represent the same concept. The IOOS community has realised this and tried to resolve it by bringing semantics into this plain SOS solution. This is described in the next section.

4 OpenIOOS: integrating ocean observations with simple semantics

This section describes the second approach to the ocean observation integration, the OpenIOOS³ portal, developed by the OOSTethys⁴ community of software developers and marine scientists. The OpenIOOS portal uses IOOS SOS services, but improves them by adding a low-level of semantics and semantic mediation.

It provides real-time data maps which show a coastal monitoring network with hundreds of in situ and remote sensors from the IOOS organisations. The real-time maps are updated by regularly polling of the OGC-compliant web services from each data provider. On the OpenIOOS portal map, each marker represents an observation available in that region. When the user clicks a marker, an information dialogue pops up showing an observation time-series. The user can filter the observations by selecting the region and/or the observation organisation. For example, the user can get information of an observation provided by the NDBC SOS. The details of this observation include the platform, the geolocation, water temperature and wind.

We now turn our attention to the semantic ingredients in the OpenIOOS. Firstly, the IOOS defines seven distinct variables which they believe to be the most important variables for their ecosystem-based ocean observing tools: sea water temperature, salinity, water level, currents, ocean color, waves and winds. Each data provider uses their own terminology to denote their own observed

³ http://www.openioos.org/real_time_data/gm_sos.html

⁴ <http://www.oostethys.org/>

properties. These names can be found from the OpenIOOS portal in the “All variables” drop-down list. We refer to these names as local variables. The local variables are mapped to the IOOS variables and the mappings are stored in the OOSTethys registry. For example, the IOOS variable “Sea.Water.Temperature” is mapped to a list of local variables such as “Temperature”, “R.TEMP”, “wassertemperature”, and so on. The adding of this simple semantic modelling improves interoperability and integration of the IOOS services. Users can use the IOOS variables to specify their queries which can be translated to queries for each SOS service through the registered mappings. Secondly, users can filter the observation results by selecting an observation region or organisation.

Though the OpenIOOS portal has greatly improved the usability of the plain OGC SOS services, its query capability is still very limited. The usage of semantics is simple and limited to the terminology level. No relationships between concepts are defined and used. More over, it only allows for discovery and retrieval of ocean observations, but has no capability to integrate the data nor to process it. Another limitation is that the information is only available to view when there is an interaction between the user and the portal and only available at the interaction point.

It is worth pointing out that the IOOS community is moving towards bringing in richer semantics into the framework. Embedding formal ontology references into OGC SWE service schemas is one future direction [8].

5 SIOOS: ocean observation integration with rich semantics

This section describes the third approach, a semantically-driven IOOS integration system (SIOOS) which also uses SWE technology but makes rich use of ontologies and mappings to offer a flexible, semantically-driven ocean observation integration platform. SIOOS is based on our Semantic Service Architecture (SSA). We start with a brief introduction to our SSA.

5.1 The Semantic Service Architecture

Our Semantic Service Architecture (SSA) is a generic information integration architecture which has been successfully applied to multiple domains such as management of water resources [4] and health research data [9, 10]. The vision of our SSA world comprises physical resources, resource ontology or resource descriptions, reference (domain) ontology, mappings between reference ontology and resource descriptions, a problem model, and a sophisticated Web services based run time environment.

An overview of the SSA architecture is depicted in Fig. 1. The Composer’s Workbench (CWB) is the front end of the SSA prototype. It provides access to layers of platform functionality through a number of specific panes: Concept Composer (visible in Fig. 2) for visual specification of semantic compositions; Mapping Rule Manager (not shown due to space limitations) for visual mapping

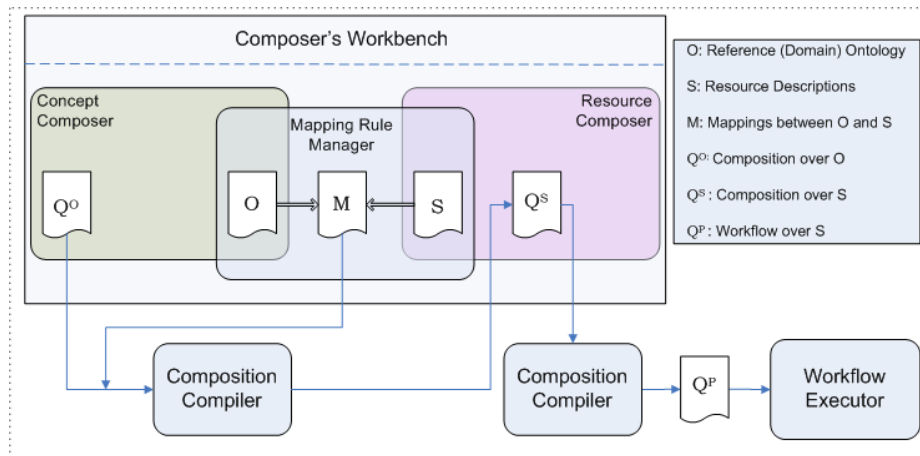


Fig. 1. The Semantic Service Architecture

specification and management of the mapping layer; and the Resource Composer (visible in Fig. 4) for visual specification of compositions of resources. The CWB also coordinates the transition of a specification between layers: conversion of a concept composition into a resource composition; and conversion of a resource composition into a physical workflow by invoking the compilation layer. And finally, submission of the workflow to the runtime.

We use O , S and M to denote the reference ontology, the resource descriptions and the mappings between the ontology and the resource descriptions, respectively. The concept composer allows users to specify *semantic compositions* using the reference ontology. We define a *semantic composition*, Q^O , as a conjunctive query expression over ontology concepts, roles and datatype properties. We define a *resource composition*, Q^S as a logical configuration of resources that can be realised as a workflow, Q^P over resources. The Mapping Rule Manager manages mappings from resource descriptions S to the reference ontology O . A first-order predicate calculus mapping language has been used to map resource descriptions to the semantic model [5]. The Resource Composer manages the compositions over ground resources such as database tables and web service operations. These kind of compositions are called *resource compositions*. The Composition Compiler translates a semantic composition Q^O to a resource composition Q^S by using a set of mappings M . It also translates a resource composition Q^S to a workflow Q^P which is executed by the Workflow Executor. The SSA allows users to specify compositions from either concept or resource level.

The SSA platform's physical layer consists of tools, client libraries or wrappers for consuming, manipulating or generating standards-based resources. Resources include WSDL web service descriptions, SOAP payloads, XML documents, XQuery programs, BPEL workflows, database tables and SQL queries. This physical layer enables the SSA platform to interact (at the level of standards

compliance) with resources external to the platform. This layer is represented in the CWB through the services pane in the Resource Composer. Some federation variability may be addressed at this level, such as different SOS versions having different WSDL wrappers.

The Resource Composer configures atomic resources drawn from the physical layer (e.g. specific parameters for a web service operation invocation); and composition of physical (e.g. chaining of services) or logical resources (a previously specified composition) into resource compositions. At this level, the SSA platform is able to configure, co-ordinate and invoke external resources. Domain experts can deal with federation variability using resource compositions. For example, CSV or XML time-series return formats are “standardized” into XML form using different XML processing chains, shown in Fig. 4.

5.2 Case studies: semantic integration over ocean observation using SIOOS

In this section, we illustrate the distinguishing features of our SIOOS system using some real-world ocean observation integration scenarios [11]. The first scenario shows the capability to integrate SOS services from different organisations with multiple observed properties. The second one is to illustrate the capability to integrate additional non-SOS services with SOS services.

Integrating SOS services and observed properties This scenario shows that users can easily and quickly query real-time ocean observations delivered by different SOS installations with multiple observed properties. This is something beyond the OpenIOOS portal’s capability.

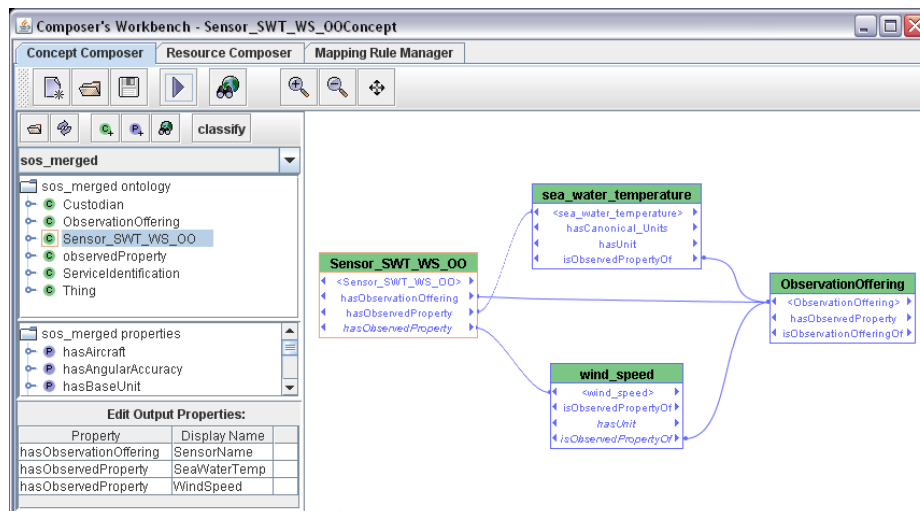


Fig. 2. A concept composition to query sensors and observations

This semantic composition uses the ontology to describe a query that seeks the names of sensor platforms and their respective measurements for both sea water temperature and wind speed. Fig. 2 shows the semantic composition Q^O formed in the Concept Composer. The *SOS_merged* domain ontology is loaded into the Concept Composer and shown in the left panel. The upper ontology panel shows the concepts and the bottom panel shows the properties. In order to form his composition, a user first creates a new concept in the ontology which is the *Sensor-SWT-WS-OO* concept here. Then the user adds properties into his new concept by dragging from the properties panel. Three properties are added into the *Sensor-SWT-WS-OO*: *hasObservationOffering*, *hasObserved-Property* and another *hasObservedProperty*. The properties of the new concept can be defined by linking them to the properties in another existing concept. Here, *hasObservationOffering* is linked to the concept of *ObservationOffering*; one *hasObservedProperty* is linked to the *sea_water_temperature* concept, and the other *hasObservedProperty* is linked to the concept of *wind_speed*.

After defining the semantic composition, the user clicks the “Run” toolbar button and the Concept Composer communicates with the composition compiler to generate a grounded resource composition (Q^S) which can be viewed. The user compiles this resource composition and a workflow (Q^P) is generated. Finally the user executes the workflow and the resulting sensor platforms and observation measurements for sea water temperature and wind speed are returned in an HTML format. Alternatively, we can automatically create and deploy a SOAP web service encompassing the generated workflow. This new web service is able to be invoked with parameter value constraints such as sea water temperature lower than 10°C.

Integrating new services with SOS services This scenario illustrates the capability provided by our SIOOS to easily and quickly integrate new services or simulation models with IOOS SOS services. In this scenario, we want to query and retrieve particular timeseries data from IOOS SOS services, process this data and plot these data on a chart. The first task is executed by the SOS services as was shown in the first example. The last two tasks are executed by two new web services that were developed by ourselves. One is a tool to process timeseries data to a standard format. The other is a tool to take the standard format data and output a chart. Mappings M are defined via the Mapping Rule Manager from SOS timeseries data processed and passed through the chart creation service to a simple OWL ontology describing the result. Next, we create a semantic composition in the Concept Composer (Fig. 3) to ask for a URL of a standard form of time series plot of sea water temperature measurements taken throughout November 2008. In this semantic composition, the concept “MySimpleGraphs”, which has only one property “hasURL”, is a new concept representing our semantic query of an URL. This URL property is linked to the “hasURL” property of the “TimeSeriesPlot” concept. The timeseries data for this “TimeSeriesPlot” is linked to the concept of “StandardFormTimeSeries”. The two properties of this concept, “isStandardFormOf” and “hasTimeInterval”, are linked to “SeaWaterTemperature” and “November 2008”, respectively.

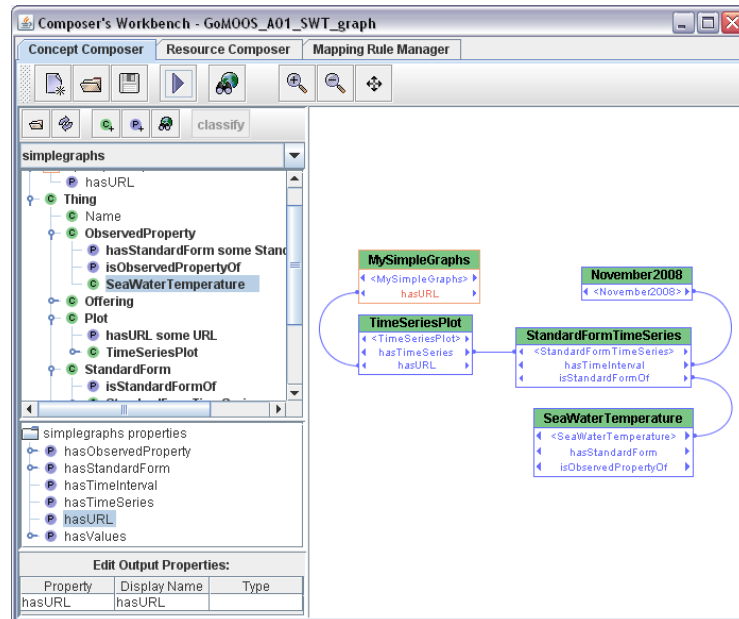


Fig. 3. A concept composition for creating charts for SOS time series

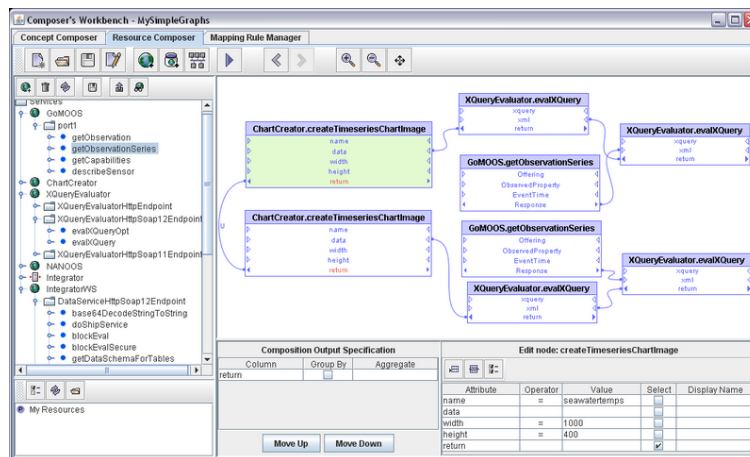


Fig. 4. The resource composition translated from the concept composition in Fig. 3

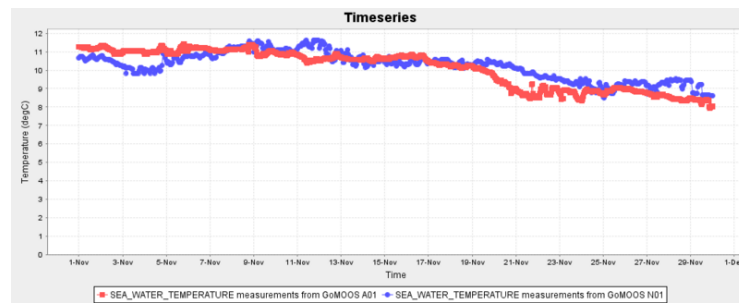


Fig. 5. A graph generated from executing the concept composition in Fig. 3

Clicking on the run button will translate this semantic composition into the resource composition of Fig. 4. Compiling this resource composition generates a workflow. Once workflow execution is complete, a timeseries graph such as Fig. 5 is available.

6 Discussion

In this section we refer back to the challenges for IOOS identified in section 2 and use that framework to analyse the three technological solutions presented.

Facilitates data access. The plain SOS approach enables common client software to access and download data from multiple service providers because of the standardised data model and interaction protocol. Because of its origins in the GIS community, the SOS services have been designed to do a very good job of uniformly representing earth-surface location of observational data.

However, experiments using Sensor Observation Services from NANOOS⁵, GoMOOS⁶ and NDBC⁷, has demonstrated the heterogeneity that exists in the federation and even in the standards-based solution. With services supporting a mixture of SOS 0.0.31 and SOS 1.0 specifications and different providers offering different return formats (some offer “standard” O&M while others use the SOS extension to only offer ioos/0.6.1) there is a non-trivial burden placed on data consumers to accommodate both federation and standards-based heterogeneity.

Probably because of the SOS origins in the GIS community, the very firm agreement on the representation of location information across services also enables very convenient, straightforward representation of observation locations from multiple services on common maps. This is exploited in the IOOS portal, but will work for any client tool designed for plain SOS.

In an independent project [12], standard OGC services were exploited for a similar data sharing problem for inland water observations in Australia. In

⁵ Northwest Association of Networked Ocean Observing Systems. <http://www.nanoos.org/>

⁶ Gulf of Maine Ocean Observing System. <http://www.gomoos.org/>

⁷ National Data Buoy Center. <http://www.ndbc.noaa.gov/>

this case a different OGC standard service was chosen to deliver the data (the OGC Web Feature Service), whilst other relevant observations were separately available via the SOS as discussed in this paper. Despite adherence to the same package of standards, the non-uniformity problem arises again. The plain SOS standardisation approach has nothing to offer for this problem, and the portal approach can only solve it by adding new Web Feature Service client capability. It is, of course, not clear how many such independent standards should be accommodated to reach a desirable coverage for a multidisciplinary community.

The SIOOS approach to data access exploits the standard, where it exists, but also accommodates variability from both the federation and the standard. Versions of the standard, as well as variations in return formats are represented and modelled in the SSA resource and mapping representations. This allows SIOOS to generate workflows that understand and respond to these uneven features of the IOOS federation.

Facilitates data discovery and understanding through contextual placement. The SOS standard does not resolve the federation problem that similar (perhaps identical), phenomena are identified by different names from different providers (e.g. “Sig_wave_Height” and “Significant_Wave_Height”). Within the SWE package of standards, this resolution role is placed on the catalogue service. The IOOS portal approach, through its mapping of terms to a less specific but common vocabulary across the federation certainly improves integration. At least it offers a more convenient interface for discovery of desirable services. However, it does not offer any method for interpretation of the remaining data, nor interpretation of that data in the problem context at hand.

A scientist seeking observations about wave heights may need to search through a catalogue service to find names for observation properties that seem to be related in the plain SOS approach. In the portal, the same scientist is certainly helped by the embedded mappings which suggest a number of observed properties that relate to wave height. On the other hand, a technician cannot use the portal to assist finding information on “system voltage”, such as the “telemetry voltage” observed property of some services.

An extension of the portal approach being proposed (see [8]) admits a range of alternative semantic mappings that can be retrieved from an external registry according to the disciplinary context of the user. Coupled with a proposed ontology reference mechanism for standardised naming of observed properties within the SOS service, this considerably expands the reach of the portal approach to a wider discipline community.

On the other hand, the SIOOS naturally supports contextual knowledge placement through three mechanisms: capture (and reuse) of resource-specific knowledge as it relates to resource or domain ontologies; mappings within ontologies; and mappings onto the scientist’s problem model.

This expands the ability to support multidisciplinary access even further. A different ontology is recommended for different identifiable user communities and the ontology provides both focal point for access and also carries a great deal of modeling context for the mappings to resources. Mappings can operate over any

aspect of the data – not just the observed property alone. Further, in contrast to the IOOS portal mapping approach above, it does not require any modification to the existing encoding of observed properties within the SOS standards.

Facilitates sound application of data through integration and interpretation. Once a scientist needs to go beyond a cursory examination of data, it is clear that tools and services that lay beyond the scope of SOS (and SWE) standards are needed. Scientists can use the IOOS portal to discover and download data, but after that, they are on their own. Having downloaded data from multiple IOOS SOS services, the scientist must contend with both federation and SOS standards unevenness. Having dealt with those issues, our scientist now needs to apply tools (such as visualization or statistical checks) to the data, and then use the (possibly processed) data and tools to address his problem.

In the examples of semantic integration, we have shown how SIOOS can be used to specify a semantic time-series processing problem—that of charting multiple sea water temperature timeseries. Further, the SSA handles a range of back-end data and service formats natively. The REST-based Web Feature Services can be accommodated in the same manner as the SOS services described here, through configuration parameters and semantic mappings. In this way, other “standard” web-services can be made available for user access and composition quite transparently. The capability for incorporation of other tools and non-standard services comes at the cost of requiring the development of the ontology and mappings to initialise the system. A user needs more ontology enabled skills to work with the system, but on the other hand, these skills also replace technical programming skills required for effective use of the data in the other two approaches.

Supports autonomy (of service provision by providers and regional solutions). The SOS approach for delivery of data inherits considerable advantages for provider autonomy over traditional data sharing mechanisms due to its adoption of a service-oriented style; these generic advantages of a service oriented approach are well documented and not repeated here. More specifically relevant for our case, while the SOS definition is not prescriptive about some aspects of data representation (for example the name, and the form of representation of the name, of an observed property) it is prescriptive about the encoding format for most of the data. Legacy data must be processed off-line by custom tools to be translated into the desired format and loaded into the server software that supports the standard service. Although it is an open standard and explicitly supports extensibility mechanisms, use of these mechanisms require extensions to the client and server tools and thereby immediately sacrifice the advantages gained by adopting standards.

The level of autonomy offered to service providers by the SOS approach is inherited directly by the IOOS portal, although its semantic mapping capability assists the user to cope with the non-standardisation of observed property terminology in plain SOS.

The SIOOS approach is able to make the best of common representation wherever it exists, so the existence of the SOS service standardisation is ex-

ploited. However, it is also capable of admitting different architectures for data delivery through its configuration capability, and different data representation formats and data models through its semantic mapping capability. A greater level of provider autonomy over these matters is thereby enabled.

Low cost (of acquisition and maintenance). Adoption of a service-oriented architecture together with a recognised data standard, as in the plain SOS approach, is a relatively low cost method to establish a large information system covering multiple agencies. It enables agencies to make their contribution independently on a best-effort basis as funding is available. Off-the-shelf client tools are also readily available at low cost. The purpose-built semantic IOOS portal is a low added cost enhancement offering service to a wider community. On top of the plain SOS cost, the SIOOS is a more expensive solution, requiring investment in developing a rich ontology for each user community and mappings to services and other resources. However it also supports integration of legacy data through (for example) a database offering or from non-SOS services, and it will sometimes be cheaper to use this capability than to require a SOS service to be installed by a provider. Because it can offer broader integration capabilities (for example by incorporating simulation tools) it might be a cost effective tool for advanced users in concert with the plain SOS provision and the simpler semantic portal offering. In particular, its ability to create new virtual services from the pre-existing services will meet a need to offer multiple service interfaces to the underlying rich information resources. Community-specific semantic modelling and mapping is the key to its broad applicability.

7 Conclusion

In this paper we have presented some of the many challenges for users and data contributors of a federated Integrated Ocean Observing System. We have reviewed three approaches aiming to address these challenges: a standards-based approach; a simple semantic terminology mapping approach; and a rich-semantics approach. The approaches build on gains made by each predecessor. We analyse the three approaches for their abilities to serve the needs of the ocean-observing community.

We find that an SOA architecture, common to all three approaches, supports autonomy of data and service provision at a relatively low cost.

We find that a standards-based approach facilitates data access through uniform interfaces and client tools, but users require a deep understanding of the meaning of data in order to apply the data and to integrate with other services from the federation.

The simple terminological mapping approach, at no cost extra to service providers, unifies the data discovery process over key properties, and so helps users to locate candidate data but does not help users understand the data or how to use it.

A rich-semantics approach can offer more to users by exploiting domain knowledge and its relationships to available resources. This aids discovery through

higher fidelity data descriptions and also enables data integration and service composition over the federated services and beyond. This capability comes at no cost to the service providers, but requires a significant initialisation task to capture the domain knowledge and to configure mappings for the services. It also demands that the user develops a sophisticated semantic literacy to use a complex ontology-enabled platform. On the other hand, such a user may create specialised services to support other users in a more conventional way.

We conclude that the latter rich-semantics approach is best at meeting the IOOS challenges, but building on the other approaches, requires a greater investment by the ocean observation community to gain the full benefit.

References

1. Duchesne, P., Maué, P., Schade, S.: Semantic annotations in OGC standards. Discussion Paper OGC 08-167, Open Geospatial Consortium (November 2008)
2. Sheth, A., Henson, C., Sahoo, S.S.: Semantic sensor web. *IEEE Internet Computing* (July/Aug 2008) 78–83
3. Bermudez, L., Bogden, P., Bridger, E., Creager, G., Forrest, D., Graybeal, J.: Toward an ocean observing system of systems. In: *OCEANS 2006*, Boston, MA, USA (Sept. 2006) 1–4
4. Ackland, R.G., Taylor, K.L., Lefort, L., Cameron, M.A., Rahman, J.: Semantic service integration for water resource management. In: *International Semantic Web Conference*. (2005) 816–828
5. Cameron, M.A., Taylor, K.L.: First-order patterns for information integration. In: *ICWE*. (2005) 173–184
6. Pattiaratchi, C.: Science and Implementation Plan for the West Australian Integrated Marine Observation System (WAIMOS). Technical report, School of Environmental Systems Engineering The University of Western Australia (2008)
7. Botts, M., Percivall, G., Reed, C., Davidson, J.: OGC Sensor Web Enablement: Overview And High Level Architecture. Technical report, OGC (December 2007)
8. Bermudez, L.E.: OGC Ocean Science Interoperability Experiment Phase 1. Technical Report 08-124r1, Open Geospatial Consortium (2009)
9. Taylor, K.L., O’Keefe, C.M., Colton, J., Baxter, R., Sparks, R., Srinivasan, U., Cameron, M.A., Lefort, L.: A service oriented architecture for a health research data network. In: *SSDBM ’04: Proceedings of the 16th International Conference on Scientific and Statistical Database Management*, Washington, DC, USA, IEEE Computer Society (2004)
10. Taylor, K., O’Keefe, C.M., Colton, J., Baxter, R., Sparks, R., Cameron, M., Lefort, L., Srinivasan, U.: A data network for health e-research. In: *Proceedings of the European Conference on eHealth (ECEH06)*. Volume 91 of *Lecture Notes in Informatics, Proceedings.*, Fribourg, Switzerland, Gesellschaft fu Informatik (GI) (October 12-13 2006) 215–226
11. Cameron, M.A., Wu, J.X., Taylor, K., Ratcliffe, D., Squire, G., Colton, J.: SIOOS: Semantically-driven Integration of Ocean Observing Systems. In: *The 8th International Semantic Web Conference (ISWC 2009) Demonstrations*, Washington DC., USA. (2009)
12. National Water Commission: Australian Water Resources 2005, Australian Water Resources Information System: System Architecture. Commonwealth of Australia (2006)