

Semantically enhancing SensorML with Controlled Vocabularies in the Marine Domain

Alexandra Kokkinaki, Louise Darroch and Justin Buck
The British Oceanographic Data Centre (BODC)
National Oceanography Centre (NOC)
Liverpool, UK
alexk@bodc.ac.uk

Simon Jirka and the Marine Profiles for OGC Sensor
Web Enablement Standards Team,
52°North GmbH,
Muenster, Germany
jirka@52north.org

Abstract—During the last decade, data produced by sensors have increased exponentially in the environmental domain. Standardization is necessary in order to integrate data originating from disparate sensor networks. Over the last few years, marine organizations and communities have been working towards the standardization of sensors, by implementing OGC SWE (Sensor Web Enablement) standards, i.e. Sensor Model Language (SensorML) to describe sensor metadata, Observations and Measurements (O&M) to describe sensor data and Sensor Observation Service (SOS) to serve them to the world. In addition, many European and US projects such as AtlantOS, SenseOCEAN, BRIDGES, XDomes, FixO3, PANGEA etc. have been implementing OGC SWE standards to achieve machine to machine communication and interoperability with other sensor networks.

SensorML is an XML based language that was purposely defined to offer many degrees of flexibility, to describe sensors with different requirements across different domains. SensorML is lenient enough to allow user generated terms to be encompassed in its syntax. As convenient as it sounds, this flexibility can result in many different variations of sensor descriptions, which reduce interoperability and discoverability via the Web. To resolve this, it is important to bring together potential user communities, identify lists of required terms, define them and then use controlled vocabularies to publish them according to standards.

In this paper, we will describe the ongoing work done by the marine community, through the Marine SWE Profiles collaboration, to create a more restrictive, semantically richer subset of SensorML, by identifying, formalizing and publishing on the Web the required terms and their definitions according to W3C standards.

Keywords— *Controlled vocabularies; XML; soft typing; SensorML standardization; NVS2.0*

I. INTRODUCTION

The marine domain has started to implement Open Geospatial Consortium's (OGC) Sensor Web Enablement (SWE) standards in order to make all types of sensors, transducers and sensor data repositories discoverable, accessible and useable via the Web. OGC SWE standards were deliberately designed to be generic enough to encompass

a wide variety of domains and disciplines that utilize sensors to make observations. Such flexibility increases the risk of slightly different implementations of the standards, which prevent sensors from becoming fully interoperable and discoverable via the Web.

The Open Geospatial Consortium's SWE activities aim at enabling "Sensor Webs", through which applications and services will be able to access sensors of all types over networks, such as the internet, and with the same standard technologies and protocols that enable the Web. These initiatives have defined, prototyped and tested several foundational components needed for a Sensor Web, namely: the Sensor Model Language (SensorML), Observations and Measurements (O&M), Sensor Planning Service (SPS), Transducer Markup Language (TML), Sensor Alert Service (SAS), Sensor Observation Service (SOS) and Web Notification Service (WNS).

In this paper we concentrate on Sensor Model Language (SensorML) which primarily focuses on providing a robust and semantically-tied means of defining processes and processing components associated with the measurement and post-measurement transformation of observations. The main objective of SensorML, which is an XML based language, is to enable interoperability, first at the syntactic level and later at the semantic level (by using ontologies and semantic mediation), so that sensors and processes can be better understood by machines, utilized automatically in complex workflows, and be easily shared between intelligent Sensor Web nodes [1]. Syntactic interoperability, which is the ability of two or more systems to communicate with each other, is solely achieved by the wide adoption of the standard by the communities that wish to communicate. Semantic interoperability, which is the ability to automatically interpret the information exchanged meaningfully and accurately, requires that both sides defer to a common information exchange reference model. The content of the information exchange requests is unambiguously defined: what is sent is the same as what is understood [2].

SensorML was designed to describe different types of domain and discipline independent processes. As its creators state, "In order to achieve interoperability within and between

various sensor communities, implementation of SensorML will require the definition of community specific semantics (within online dictionaries or ontologies) that can be utilized within the framework” [3]. This is true since most properties in SensorML utilize the concept of "soft-typing". That is, rather than trying to pre-define in the schema every possible property that might be used to describe a particular sensor or might be measured by a sensor, SensorML allows property types to be defined outside of the SensorML schema (typically within an online ontology) and then be used within SensorML as a value to the definition attribute. The value of the definition attribute must be a resolvable URL that references an online property definition or single entry within an ontology [4].

The list of properties that can be used to describe a sensor can grow long especially if more than one sensor domain is considered. The following example demonstrates the difficulties encountered when using SensorML to describe a sensor. In this example, User A classified “fluorometer X” under the category of “Fluorometers”. As shown in Figure 1, User A defined a property named “Instrument Type” and published it in an ontology. He then added the free text value “Fluorometer” as the value of the instrument type. User B, following the same pattern, chose the term “Sensor Category” to classify his sensor and assigned the term “active flurometer” - wrongly spelled - as a value to the property. Software client X failed to discover all available fluorometers, as it used different term definitions and term values from those used by User A and User B.

```
<smil:classification>
  <smil:ClassifierList>
    <smil:classifier>
      <smil:Term definition=
        "http://www.example.com/definitions/Instrument Type/">
      <smil:label>Instrument Type</smil:label>
      <smil:value>Fluorometer
    </smil:value>
    </smil:Term>
  </smil:classifier>
</smil:ClassifierList>
```

Fig. 1. User A SensorML description

```
<smil:classification>
  <smil:ClassifierList>
    <smil:classifier>
      <smil:Term definition=
        "http://www.example2.com/definitions/Sensor Category Type/">
      <smil:label>Sensor Category</smil:label>
      <smil:value>Active flurometer
    </smil:value>
    </smil:Term>
  </smil:classifier>
</smil:ClassifierList>
```

Fig. 2. User B SensorML description

SensorML creators have identified the benefit of ontologies since the publication of the standards by stating that: “Sensor ontologies are becoming increasingly important for creating standard dictionaries of sensor-related terminology and for mapping relationships between these terms.” Many sensor technologies, including the Sensor Web Enablement (SWE) encodings and Web services, depend on and benefit greatly from the existence of online, resolvable ontologies of terms related to sensors. SensorML creators have

created the SensorML ontology to list these terms, through the Marine Metadata Interoperability (MMI) project. It hosts an Ontology Registry and Repository hosting a number of small project specific controlled vocabularies [5]. Since different communities require different terminologies, the ontology can fulfill only a subset of the required concepts.

The marine community has been using controlled vocabularies, i.e. standardized sets of terms, in tagging metadata and labeling data in order to solve the problem of ambiguities associated with data markup and enable records which are interpreted by computers. Controlled vocabularies for the marine community are served by a number of servers including the NERC Vocabulary Server version 2.0 (NVS2.0). This server provides access to lists of standardized terms that cover a broad spectrum of disciplines relevant to the oceanographic and wider community. NVS2.0 makes use of the World Wide Web Consortium's Simple Knowledge Organization System (SKOS) to represent knowledge in a format understandable by computers. In SKOS, vocabularies are modeled as collections and terms are modeled as concepts. Collections and concepts have unique URIs that are resolvable through a RESTful interface to either HTML or RDF documents through content negotiation. Collections are also accessible through SOAP Web Services and a SPARQL endpoint¹.

In this paper we present an initiative from a collaboration within the marine community to create and maintain several controlled vocabularies, to semantically enhance SensorML and bring semantic interoperability amongst environmental sensor networks.

II. METHODS

Our work to semantically enhance SensorML for the marine domain comprises four distinct steps. Step one is the formalization of the concepts and definitions used to describe sensors in the marine domain and their organization in collections. The next step is the publication of these concepts and collections using unique URIs. Step three is the definition of internal mappings between concepts and other NVS2.0 concepts sharing the same meaning. The last step is the definition of external mappings with overlapping concepts from the SensorML ontology and other well-known vocabularies, e.g. DBpedia, thus making sensors more accessible and discoverable via the Web.

A. Marine SWE Profiles

To avoid interoperability issues in OGC SWE implementations by different organizations and users, an agreement was needed on how to apply SWE concepts and how to use vocabularies in a common way that would be shared by different projects, implementations, and users.

Partners from several projects and initiatives (AODN, BRIDGES, ENVRI+, EUROFLEETS /EUROFLEETS2, FixO3, FRAM, IOOS, Jerico/Jerico-Next, NeXOS, ODIP/ODIP II, RITMARE, SeaDataNet, SenseOcean, X-

¹ <http://vocab.nerc.ac.uk/sparql/>

DOMES) created the Marine SWE Profiles group as a solution to the need mentioned above. They joined forces to develop common marine profiles of OGC SWE standards that can be used in multiple projects and organizations. [6]

Marine SWE Profile members interact and communicate through the use of a mailing list and a wiki website. The wiki helps to collect and discuss different approaches to how OGC Sensor Web Enablement (SWE) standards (Sensor Observation Service (SOS), Observations and Measurements (O&M) and SensorML are used in different projects and systems. It is currently structured in the following subsections, which can be edited by its members after logging in:

- **SweExamples:** Examples of SensorML, O&M and SOS usage
- **SweVocabularies:** Vocabularies for the Marine SWE Profiles
- **SweProfile:** Structure and proposed content of the Marine SWE Profiles
- **SosInventory:** Inventory of SOS Servers

The Marine SWE Profiles mailing list is essentially a discussion list. Members are allowed to post their own items which are broadcast to all of the other mailing list members. For the purposes of vocabulary building, the list was given the responsibility to act as the SensorML vocabulary content governance, which is important in order to stay up-to-date and in sync with ongoing developments.

The publication of SensorML implementations by different projects revealed the lack of published vocabularies for term and property definitions and the need for common vocabularies to refer to the same terms coherently in the marine domain.

B. NVS2.0

The NERC Vocabulary Server version 2.0 (NVS2.0) provides access to lists of standardized terms that cover a broad spectrum of disciplines relevant to the oceanographic and wider community.

NVS2.0 is based on the Simple Knowledge Organization System (SKOS) model. SKOS is based on the "concept", which it defines as a "unit of thought", that is an idea or notion. In NVS2.0, each vocabulary is a collection and owns a unique URI that resolves, after content negotiation, in a self-descriptive RDF document or an HTML page if a machine or a human entity requests it respectively [7].

NVS2.0 URIs are published using the following pattern: <http://vocab.nerc.ac.uk/collection/XXX/current/> for collections and <http://vocab.nerc.ac.uk/collection/XXX/current/YYYYYY> for concepts, where XXX is a three character code referring to a vocabulary collection and YYYYYY is a variable length code uniquely identifying each concept in the collection, e.g. <http://vocab.nerc.ac.uk/collection/P07/current/3AKCHY57/>.

Each controlled vocabulary delivered by NVS2.0 contains the following information:

- Key — a compact permanent identifier for the collection, designed for computer storage rather than human readability
- Title — a text string representing the title of the vocabulary in human-readable form
- Abbreviation — a concise text string representing the title in human-readable form where space is limited
- Date — latest publication date
- Definition — full description of what the vocabulary describes.
- Creator — the organization that created the vocabulary
- Owner — the organization that owns the vocabulary
- Manager — the organization that manages the vocabulary
- Publisher — the organization that publishes the vocabulary

The RDF snippet in Figure 3, demonstrates the information originating from W05² vocabulary that contains SensorML characteristic terms.

```
<skos:Collection rdf:about
"http://vocab.nerc.ac.uk/collection/W05/current/">
<skos:prefLabel>SensorML Characteristic Section Terms
</skos:prefLabel>
<dc:title>SensorML Characteristic Section Terms</dc:title>
<skos:altLabel>SensorML Characteristics</skos:altLabel>
<dc:alternative>SensorML Characteristics</dc:alternative>
<dc:description>Terms used in SensorML to describe
properties of an observation system that do not further qualify
or quantify its output values.</dc:description>
<dc:date>2016-09-15 02:00:04.0</dc:date>
<owl:versionInfo>2</owl:versionInfo>
<grg:RE_RegisterManager>British Oceanographic Data
Centre</grg:RE_RegisterManager>
<dc:publisher>Natural Environment Research Council
</dc:publisher>
<dc:creator>Sensor Web Enablement Marine Profiles
</dc:creator>
<grg:RE_RegisterOwner>Sensor Web Enablement Marine
Profiles</grg:RE_RegisterOwner>
<rdfs:comment>Governance for vocabularies created for use
in SWE Marine Profiles.</rdfs:comment>
</skos:Collection>
```

Fig. 3. RDF code for NVS2.0 vocabularies

Additionally, vocabularies contain lists of terms classified as SKOS concepts, each one having a unique URI resolving to an RDF or HTML document, as for collections. The controlled

² <http://vocab.nerc.ac.uk/collection/W05/current/>

vocabularies delivered by NVS2.0 contain the following information for each term:

- Key — a compact permanent identifier for the term, designed for computer storage rather than human readability
- Term — the text string representing the term in human-readable form
- Abbreviation — a concise text string representing the term in human-readable form where space is limited
- Definition — a full description of what is meant by the term

All of the vocabularies are fully versioned and a permanent record is kept of all changes made. NVS2.0 can be accessed in three different ways: through a SOAP service, a RESTful interface and a SPARQL endpoint.

NVS2.0 was chosen by the Marine SWE Profiles community to publish SensorML terms, as it and its predecessors have successfully served the marine community for more than ten years. The use of NVS2.0 within the European Union SeaDataNet project is outlined in [8]. In the wider arena, the Ocean Data Interoperability Platform (ODIP) is an international collaboration of data management organizations which includes SeaDataNet. They are fostering best practices and common standards. In addition, they are creating prototypes to enable the transfer of technologies. The NVS2.0 has been utilized within ODIP prototype 2 to underpin interoperability by linking EU, US and Australian research cruise programs by providing cruise information at an international level.

III. RESULTS

This work, which is based on standards, aims to semantically enhance SensorML in the marine domain according to W3C standards. Thus, it allows computers not only to communicate, but also to seamlessly understand the communicated information.

There are essentially two sections in SensorML that would benefit by the use of vocabularies: The term definition and the term value, as shown in Figure 4.

For “term values”, Marine SWE Profiles members agreed to use existing concepts in NVS2.0. The following collections were identified to adequately serve term values:

- Observable property: NVS2.0 Collections P01, P07
- Instrument Type: NVS2.0 Collection L05
- Platform Type: NVS2.0 Collection L06
- Roles: NVS2.0 Collections G04, C86
- Feature of Interest: NVS2.0 Collection C19
- Manufacturer: NVS2.0 Collections L35, C75

NVS2 Collection P01, which lists terms used to describe individual measured phenomena and P07, which is a list of the

Climate and Forecast standard names, have been nominated to serve SensorML observable properties. The L05 collection,

```

<sm:classification>
  <sm:ClassifierList>
    <sm:classifier>
      <sm:Term definition=
        "http://vocab.nerc.ac.uk/collection/W06/current/CLSS0002/">
      <sm:label>Instrument Type</sm:label>
      <sm:value>http://vocab.nerc.ac.uk/collection/L05/current/113/
    </sm:value>
    </sm:Term>
  </sm:classifier>
</sm:ClassifierList>
</sm:classification>

```

Fig. 4. SensorML code snippet

which lists device categories, is used for the classification of instruments and procedures. L06, in the same respect, provides a list of platform categories to be used for classifying platforms. G04 and C86 list roles and populate SensorML’s role property. C19, which is the Salt and Fresh Water Body Gazetteer, can be used to create a rich list of features of interest. L35 and C75 can be both used to populate the manufacturer property, since they refer to organizations and manufacturers respectively

The absence of a standard list of term definitions initiated the SWE Marine members’ collaboration and agreement. The SWE Examples wiki subsection was used to collate the various SensorML descriptions posted by the members. Subsequently, the group identified the common terms under each section and provided a common name for the semantically same but differently named terms. The terms were also enhanced with definitions and alternative labels and were published on the SWE Vocabularies wiki page for review and final approval. The list was then submitted to the Vocabulary Management Group at BODC. They performed final checks on integrity and conformity before accepting the list for publication.

For new terms and vocabularies, a new process has been established. Members are encouraged to post the desired set of terms on the wiki, complementing it with a title and a definition. The set is then checked by the Vocabulary Group in BODC and if any changes are applied, it is posted again on the wiki. The group needs to approve the changes to finally be published on the Web. Disagreements are discussed in the mailing list.

SensorML consists of sections which include several terms. In NVS2.0, each section is modeled as a new vocabulary, holding a unique URI, listing a set of domain relevant terms. Following the NVS2.0 URL pattern, SensorML vocabularies are all grouped under the ‘WOX’ notation as shown in Table 1, although there is no semantic relevance between the vocabulary’s subject and the notation. Each vocabulary is self-documented and refers to the Marine SWE Profiles group as its creator and owner. BODC is the manager and moderator and NERC is the publisher.

The XML code snippet in Figure 4 displays the standardized version of the examples shown in Figure 1 and Figure 2 respectively. The different term definitions and

values were merged under unique URIs, which are accommodated in the SensorML code.

A. Mappings

As stated previously, links from NVS2.0 concepts to other data sources can only benefit metadata tagged with these concepts as they become more discoverable on the Web. The mapping process is still ongoing and the objective is to initially use the owl:sameAs property for stating that another data source also provides information about a specific NVS2.0 concept. The RDF links will be set manually for the mappings of NVS2.0 to MMI and to other NVS2.0 concepts.

Table 1. Table listing the URI and the description of the published SensorML collections

URI	Title
http://vocab.nerc.ac.uk/collection/W03/current/	SensorML History Event Types
http://vocab.nerc.ac.uk/collection/W04/current/	SensorML Capability Section Terms
http://vocab.nerc.ac.uk/collection/W05/current/	SensorML Characteristic Section Terms
http://vocab.nerc.ac.uk/collection/W06/current/	SensorML Classification Section Terms
http://vocab.nerc.ac.uk/collection/W07/current/	SensorML Identification Section Terms
http://vocab.nerc.ac.uk/collection/W08/current/	SensorML Contact Section Terms

B. Applications

Enhancing SensorML with standardized lists of terms ensures interoperability between different implementations of OGC SWE sensor descriptions. Providing these vocabularies as allowed values through drop down lists for “term values” in SensorML editors leads to interoperable SensorML descriptions. A worthy example is the EDI Metadata Editor, which is a template-driven metadata authoring tool that can be easily customized to any XML-based metadata format (e.g. SensorML) and to a specific workgroup, institute, or project. It also connects to the NVS2.0 SPARQL endpoint to provide lists of allowed terms for property values [9].

Additionally, SOS clients can leverage the standardization of SensorML to easily discover sensors based on their characteristics. For example, client software searching for Instrument Types - as defined in NVS2.0 concept <http://vocab.nerc.ac.uk/collection/W06/current/CLSS0002/> - being fluorometers - as defined in <http://vocab.nerc.ac.uk/collection/L05/current/113/> - will be able to locate all relevant sensor descriptions that have been described with the vocabularies mentioned above. As a consequence, sensors described in SensorML become standardized, more discoverable and usable via the Web.

IV. DISCUSSION AND CONCLUSIONS

The need for controlled and defined vocabularies in SensorML has been clear since its creation and this became evident as its use matured. In the marine community, the exposure of different SensorML implementations under the collaborative environment of the Marine SWE Profiles wiki and the success of NVS2.0 were the two key factors that resulted in the creation of the SensorML vocabularies. As stated in [10], vocabularies should be published by a trusted group and they should be accessible for a long period. NVS2.0 fully meets these conditions. A critical element in this work was the vocabulary governance, applied by the SWE Marine Profile group, as a list of specialized users, as opposed to one authority, resulting in trust and acceptability of the new vocabularies.

Although SensorML ontology published under the MMI project offers several terms and definitions, it does not capture all of the marine domain. As a result, the SWE Marine Profile community chose NVS2.0 to host new domain-specific vocabularies. As stated in [10], mappings will be established between vocabularies where there are overlapping terms to enhance the discoverability of sensor metadata and to inform users how terms relate with each other.

The creation of the SensorML vocabularies draws the required boundaries for the uniform use of the language in the marine domain, but it also enhances SensorML semantically.

V. CONCLUSIONS

SensorML’s flexibility, specifically the soft typing characteristic, causes variability in published sensor descriptions, thereby reducing interoperability and discovery via the Web. To address this issue, Marine SWE Profiles group decided to formalize the required terms and publish them in the form of controlled vocabularies served by NVS2.0 vocabulary server. The collections and terms are governed by the group and maintained by BODC so they are assured and accepted by the community. The work is ongoing and includes mappings between terms that share common meaning in NVS2.0, SensorML ontology and other existing vocabularies.

This work, which is highly collaborative, shows what can be achieved when people reuse existing well-functioning infrastructures and join forces to handle interoperability issues.

ACKNOWLEDGMENT

This work is funded by the European projects SenseOCEAN and BRIDGES and supported by the National Environmental Research Council (NERC). SenseOCEAN is a Collaborative Project funded by the European Union Seventh Framework Programme (FP7/2007–2013) under grant agreement No. 61414. The BRIDGES project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 635359. The NVS2.0 server is supported by NERC National Capability (NC) funding for NC-services, facilities and data (NC-SFD).

REFERENCES

- [1] "Sensor Model Language (SensorML) | OGC", Opegeospatial.org, 2016. [Online]. Available: <http://www.opengeospatial.org/standards/sensorml>. [Accessed: 31- Jul- 2016]
- [2] "Semantic interoperability of health information", En13606.org, 2016. [Online]. Available: <http://www.en13606.org/the-ceniso-en13606-standard/semantic-interoperability>. [Accessed: 31- Jul- 2016]
- [3] Portal.opengeospatial.org, 2016. [Online]. Available: http://portal.opengeospatial.org/files/?artifact_id=21273. [Accessed: 31- Jul- 2016]
- [4] "SensorML 2.0 Metadata - Identifiers and Classifiers", Sensorml.com, 2016. [Online]. Available: <http://www.sensorml.com/sensorML-2.0/examples/identifiers.html>. [Accessed: 31- Jul- 2016]
- [5] J. Graybeal, A. Isenor and C. Rueda, "Semantic mediation of vocabularies for ocean observing systems", *Computers & Geosciences*, vol. 40, pp. 120-131, 2012..
- [6] S. Jirka, "Marine Profiles for OGC Sensor Web Enablement Standard", in *EGU General Assembly*, Vienna, 2016, p. 14690.
- [7] A. Leadbetter, *The Semantic Web in Earth and Space Science. Current Status and Future Directions Part II Chapter 2 Linked Ocean Data*, 1st ed. 2016, pp. 11-31 [Online]. Available: <https://books.google.co.uk/books?isbn=161499501X>. [Accessed: 31- Jul- 2016]
- [8] D. Schaap and R. Lowry, "SeaDataNet – Pan-European infrastructure for marine and ocean data management: unified access to distributed data sets", *International Journal of Digital Earth*, vol. 3, no. 1, pp. 50-69, 2010.
- [9] C. Fugazza, A. Oggioni, M. Pepe, F. Pavesi, P. Carrara. *DATA 2014 - 3rd International Conference on Data Management Technologies and Applications*. DOI: 10.5220/0004997603490356
- [10] "Best Practices for Publishing Linked Data", W3.org, 2014. [Online]. Available: <https://www.w3.org/TR/ld-bp/>. [Accessed: 31- Jul- 2016]