Integrating Oceanographic Sensor Data Using SSN/SOSA Ontology

Gustavo Nuñez[1,⋆], Carlos Buckle[1][0000−0003−0722−0949], and Marcos Zárate[1,2][0000−0001−8851−8602]

[1] Laboratorio de Investigaciones en Informática, Facultad de Ingeniería, Universidad Nacional de la Patagonia San Juan Bosco (LINVI-UNPSJB), Puerto Madryn, Argentina.
gnunez@ing.unp.edu.arcbuckle@unpata.edu.ar
[2] Centre for the Study of Marine Systems, Patagonian National Research Centre (CESIMAR-CENPAT-CONICET), Puerto Madryn, Argentina.
zarate@cenpat-conicet.gob.ar

Abstract As the deployment of ocean sensors continues to grow, there is a growing need for standardized ways to represent and integrate sensor data from different sources. One approach to achieving this is through the use of Semantic Sensor Network (SSN/SOSA) ontology, which provides a common vocabulary and framework for describing sensors, observations, and their properties. In this paper, we present a method for converting ocean sensor data in CSV format to Resource Description Framework (RDF) using RDFLib library for Phyton and SSN/SOSA ontology. The resulting RDF triples can be stored in a triplestore for querying and analysis, providing a standardized representation of ocean sensor data that can be easily integrated with other data sources.

Keywords: Ocean sensors · SSN/SOSA ontology · RDFLib · Aqualink.

1 Introduction and motivation

The ocean is a complex and dynamic system that plays a critical role in the Earth's climate and ecosystem [1]. Understanding the ocean requires the collection and analysis of vast amounts of data from a variety of sources, including oceanographic sensors. However, integrating and interpreting these data can be challenging due to the diverse nature of sensors and the lack of standardized approaches for describing them. The use of Semantic Sensor Network (SSN/SOSA) ontology [2] in ocean sciences has gained significant attention in recent years. This ontology provides a standardized framework for describing sensors, observations, and their metadata, allowing for seamless integration and interoperability of data from various sources. SSN/SOSA is designed to enable machine-readable descriptions of sensors, their capabilities, and the characteristics of the observations they make, which can be used to facilitate data discovery, integration, and interpretation.

⋆ Corresponding author.

In ocean sciences has been applied to a wide range of applications, including oceanographic monitoring, marine biology, and climate science. For example, in oceanographic monitoring, SSN/SOSA have been used to describe the properties of oceanographic sensors, such as their measurement range, accuracy, and resolution, as well as their deployment location and time series data. This has allowed for the integration of data from various sensors, including ocean gliders, buoys, and satellites, to better understand oceanographic processes and phenomena.

In climate science have been used to describe sensors used to monitor environmental parameters that affect the Earth's climate, such as temperature, salinity, and sea level. This has allowed for the integration of data from various sources, such as ships, buoys, and satellites, to better understand the Earth's climate system and its response to changes in the environment. Despite the potential benefits of using SSN/SOSA in ocean sciences, there are still challenges that need to be addressed. One of the main challenges is the lack of a standardized approach for implementing SSN/SOSA, which can lead to inconsistencies in the way sensors and observations are described. Additionally, there is a need for improved data management and data sharing practices to ensure that SSN/SOSA data is easily accessible and can be used by a wide range of stakeholders. Overall, the use of SSN/SOSA in ocean sciences has the potential to significantly improve our understanding of the oceans and their role in the Earth's system. As such, there is a need for continued efforts to develop and promote the use of SSN/SOSA in ocean science research and applications.

In this short paper, we present a method for converting ocean sensor data in CSV format extracted from Aqualink project to RDF using RDFLib [3] and SSN/SOSA ontology. A set of queries is also provided to retrieve the information from the resulting graph. The remainder of this paper is structured as follows: Section 2 presents the most relevant characteristics of the sensors used to create the graph. Section 3 shows the steps needed to convert to RDF using RDFLib and a set of queries that demonstrate its use. Finally, in Section 4, we present some conclusions related to previous experiences.

2 Background

The Aqualink Monitoring for marine ecosystems project[3] is a collaborative effort between multiple research institutions and government agencies aimed at developing and deploying an advanced monitoring system for marine ecosystems. The project is focused on the development of a standardized sensor network infrastructure that can be used to collect, integrate, and disseminate data from a wide variety of sensors, including those measuring physical, chemical, and biological parameters.

To carry out the experiments, information provided by the sensors that belong to the oceanographic buoy located in Isla Leones, Argentina (LAT: -45.043 LONG: -65.607)[4]. The station that is equipped with a suite of sensors, including

[3] https://aqualink.org/

[4] https://aqualink.org/sites/1136

those for measuring water temperature, wind speed and significant wave height. Figure 1 shows the schematic diagram of the buoy and the sensors associated with it. The data collected by these sensors is transmitted in near real-time to a central data management system, where it is quality-checked, processed, and made available to researchers and resource managers through a web-based portal. The data collected by the station can be used to inform a range of research and management activities, including the assessment of marine biodiversity, the detection of harmful algal blooms, and the monitoring of oceanographic processes such as up-welling and water column stratification.

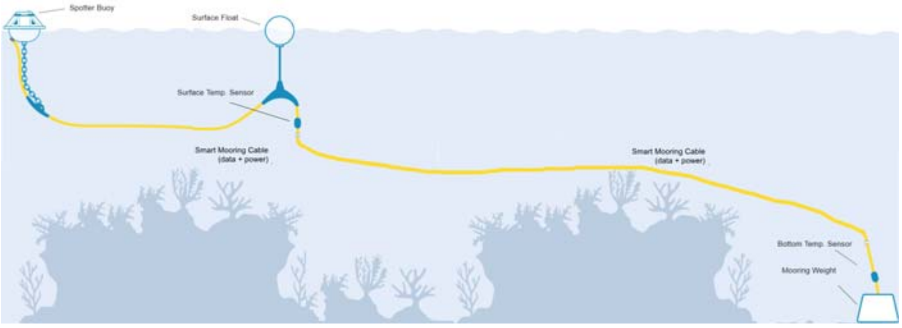

Fig. 1. Schematic diagram of the buoy and its sensors provided by Aqualink.

3 Case study

To illustrate the use of SSN/SOSA ontology in ocean sciences, a case study of Aqualink initiative is presented. The oceanographic variables that will be used to create the graph are the following: top temperature, bottom temperature, wind speed and significant wave height. The range of dates used goes from January 24, 2023 to February 24, 2023 with a total of 2418 measurements of the variables mentioned above.

We mapped the sensor properties and metadata to SSN/SOSA ontology classes and properties, including the following: `sosa:Platform`: represents a physical or virtual entity that carries one or more sensors and can act as a host for observing and measuring the environment. `sosa:Sensor`: represents a physical device that can measure, observe, or estimate properties of the environment. `sosa:FeatureOfInterest`: represents the entity or phenomenon being observed or measured by a sensor. It can be a point or an area in space, a sample or a specimen, or an event or a process. `sosa:Observation`: represents a record of an act of observing or measuring a property of a feature of interest by a sensor. It captures the values of one or more observable properties at a particular time and place. `sosa:Result`: represents the value or values of an observable property resulting from an observation and `sosa:ObservableProperty`: represents a property or phenomenon that can be observed or measured by a sensor. It can be a physical, chemical, biological, or environmental property, among others.

To model temporal entities the W3C time ontology [4] was used, to spatial entities we chose the GeoSPARQL ontology [5] and to model measurement units the Quantities, Units, Dimensions, and Types (QUDT) Ontology [6].

Figure 2 shows how station in Punta Leones was modeled. It can be seen that the buoy is of type `sosa:Platform`, the temperature sensor is of type `sosa:Sensor`, the surface temperature observations are of type `sosa:Observation`. Observation results are of type `sosa:Result` and `qudt:QuantityValue`.

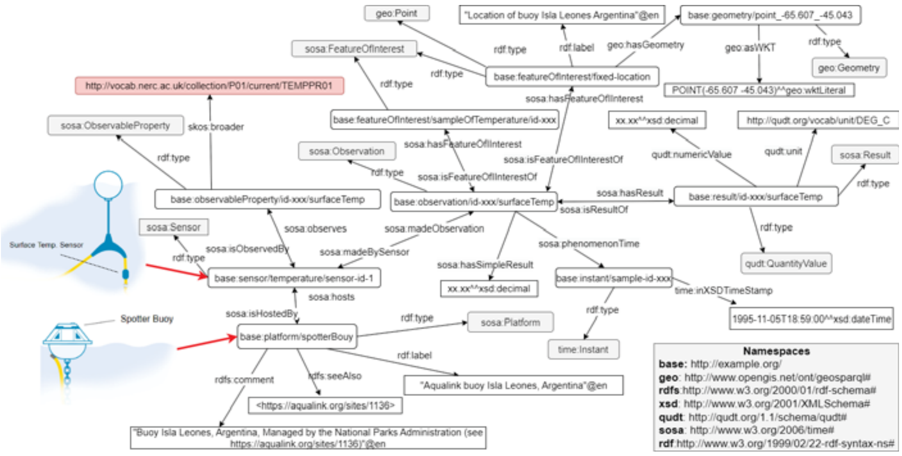

Fig. 2. Simplified diagram of the resulting graph, in the image you can see the main classes and their relationships. For simplicity the diagram only shows the modeling of the sensor that records the surface temperature..

To facilitate the integration and interoperability of data from Aqualink sensors, we converted data in CSV format to RDF using RDFlib. The approach used consists of the following steps:

1. Load CSV data into DataFrame.
2. Create RDF graph with SSN/SOSA ontology and complementary ontologies.
3. Map CSV columns to RDF properties according to Figure 2 using RDFLib.
4. Save RDF graph to file.
5. Query the graph.

Google colab was used to perform all mappings, the code can be tested in the following link. The resulting RDF data can be easily queried and integrated with other RDF data sources, enabling more comprehensive analysis of oceanographic data.

3.1 Quering with SPARQL

RDFLib can create and manipulate RDF graphs and can be used to extract information from these graphs using SPARQL query language [7].

In this case study, it is necessary to obtain surface and bottom temperatures recorded during the date range from 2023-01-28 to 2023-01-30. To do this, it is first necessary to create an RDF graph, prepare a SPARQL query that selects all triples in the graph whose type is `sosa:Observation`, and execute the query using the query() method of the Graph object. The results of the query are returned as rows, which we can iterate over and print. To see the complete listo of queries developed see the following link to google colab.

4 Conclusions

The case study presented in this paper highlights the successful implementation of SSN/SOSA in the Aqualink project, which has enabled the collection and sharing of high-quality data on a range of physical, chemical, and biological parameters in marine environments. This has the potential to improve our understanding of marine ecosystems and inform more effective management practices. Overall, the use of SSN/SOSA ontologies in oceanographic sensor data sharing and collaboration has the potential to drive significant advancements in our understanding of the ocean and the impacts of human activities on marine ecosystems. Continued efforts to develop and apply these standards will be critical in achieving more effective management and conservation of our oceans.

References

1. Elvira S Poloczanska, Christopher J Brown, William J Sydeman, Wolfgang Kiessling, David S Schoeman, Pippa J Moore, Keith Brander, John F Bruno, Lauren B Buckley, Michael T Burrows, et al. Global imprint of climate change on marine life. *Nature Climate Change*, 3(10):919–925, 2013.
2. Armin Haller, Krzysztof Janowicz, Simon JD Cox, Maxime Lefrançois, Kerry Taylor, Danh Le Phuoc, Joshua Lieberman, Raúl García-Castro, Rob Atkinson, and Claus Stadler. The modular ssn ontology: A joint w3c and ogc standard specifying the semantics of sensors, observations, sampling, and actuation. *Semantic Web*, 10(1):9–32, 2019.
3. Carl Boettiger. *rdflib: A high level wrapper around the redland package for common rdf applications*, 2018.
4. Simon Cox and Chris Little. Time ontology in owl, w3c candidate recommendation 26 march 2020. [Online; accessed 14-Apr-2023].
5. Matthew Perry and John Herring. Ogc geosparql-a geographic query language for rdf data. *OGC implementation standard*, 2012.
6. Ralph Hodgson, Paul J Keller, Jack Hodges, and Jack Spivak. Qudt-quantities, units, dimensions and data types ontologies, 2014. [Online; accessed 14-Apr-2023].
7. Sparql 1.1 query language w3c recommendation 21 march 2013, 2013. [Online; accessed 14-Apr-2023].